UNIVERSITÀ DEGLI STUDI DI TRENTO

CIMeC - Center for Mind/Brain Sciences

Doctoral school in Cognitive and Brain Sciences
XXVII cycle

Ph.D. Dissertation

Characterizing the spatiotemporal profile and the level of abstractness of action representations: neural decoding of magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI) data

Raffaele Tucciarelli

Advisor: Angelika Lingnau

A Eli

Acknowledgements

Thanks Angelika, I could not have done this work without you. Your positivity, enthusiasm and willpower inspired me. Thanks for being so open minded, so collaborative, and so helpful. I learnt a lot from you. It was great doing my PhD under your supervision.

Thanks Nathan W., for your support and encouragement, and for all the opportunities that you gave me. Thanks Luca, I learned a lot from you as well, you are a good friend. Thanks Moritz, for all your suggestions and all the “confounds” that you report.

Thanks Magda, not just for your help, for your beautiful thoughts. Thanks James, for all the fantastic and crazy conversations we have.

Thanks Marco, my brother, one of few person that knows how to make me laugh. I wish I could see you more often. Thanks Rossana, my sister, I learnt from you that it is possible to never get tired.

Thanks mom, I do not know why you believe in me so much, but it is very comforting. Seeing you being up all night for studying was useful. Thanks dad, for being so good and understanding. I think you knew better than me that I could make it through all of this.

Thanks Dima, Fasto, Fili, Fotino, Giu, Savio. I barely see you, but I feel like you are always around.

Thanks Dinos, it was great meeting you in Rovereto.

And thanks a lot, Eli. I have never met a person like you, I don’t think there are many. With your witty and clever thoughts, with your obstinacy, you have inspired me since the first day I saw you, and you keep doing it every day. You are unique. Don’t ever change.

Goodbye Trentino. It was nice, in spite of everything.

TABLE OF CONTENTS

List of tables

List of figures

Chapter 1: “What comes first? The understanding or the simulation?”

Introduction

Italians love coffee and drink several cups a day. Imagine you are Italian and you are visiting an Italian friend of yours. Most likely, he has a moka pot, a typical Italian coffee maker. When you see him holding the moka pot, you immediately understand his intention: he is about to make coffee. You make coffee yourself every day; it is easy for you to understand your friend’s intention, just by observing this simple grasping movement. He opens the moka, fills the bottom with water (not too much!), puts the ground coffee in the basket, places the basket in the pot, closes the moka and puts it on the stove. After about 5 minutes, the kitchen will be filled with the aroma of coffee. Now, consider another situation. Your friend has just bought a fancy new coffee maker, like the one that you always see at the cafe. It makes espresso. You have never made an espresso yourself, but you have seen the barista make it many times. Your friend knows all the steps needed to prepare espresso. You watch him put the ground coffee in the holder, tamp down the coffee, switch the machine on for 1 second just to remove the dirty water, twist the holder into place on the machine with a rotational movement, place the cups below the spout, and switch the machine on again. You know that after about 5 seconds the dark liquid will pour into the cups. Finally, consider yet another situation. Your friend has just visited France and he bought a French press to make French coffee. You know nothing about it. You see him holding the press, but you have no idea of what he is doing. You cannot even say that he is making coffee. Then, you see him putting the ground coffee in the glass container, now you understand what he is doing. He presses the plunger and after 4 minutes, the coffee is ready.

In all three scenarios described above, there is a common aspect: your friend’s more distal action (or intention) is to make coffee. The three ways of preparing the coffee are very different, but ultimately the goal is the same. This example shows how the same goal can be achieved by many different actions. Importantly, you can understand your friend’s intention because you can understand his actions. Different sources of information can be used to reach this goal. In the first scenario, you know how to perform those actions yourself (with the moka pot); in the second scenario, you have collected many visual experiences (watching espresso being made using a machine); in the last scenario, you have knowledge of the objects related to the action (ground coffee) that helps you to make sense of what you see. Understanding others’ actions is a fundamental human ability, especially in social situations (maybe you do not want a coffee, thus you can stop your friend from preparing it when you see him picking up the moka pot). An intriguing and still open question in neuroscience is how the brain allows us to understand situations like this. In other words, how is the brain able to relate many different ways of performing an action to one specific goal or intention (the many-to-one mapping problem)?

During my PhD, I have investigated the neural bases of action understanding. This field of research has attracted a lot of interest in the last two decades both within the scientific community and in the media. There are at least two kinds of reasons why this is the case. First, there is the theoretical reason. The main aim of psychology and neuroscience is to understand how the mind and the brain work. Low-level processes, such as early vision and simple movements, have been described extensively and we have a fairly good understanding of their neural mechanisms. For example, it has been known since the 1950s that there are areas in the central part of the brain that will produce overt movements (Penfield & Boldrey, 1937) if electrically stimulated, and that oriented edges and gratings induce neurons in primary visual cortex to discharge (Hubel & Wiesel, 1959). On the other hand, higher-level processes such as action understanding, meaning the ability to discriminate and recognize observed actions, are more complex and still poorly understood. What are the neural bases of action

understanding? This is a still an open question and finding an answer is fundamental for building a theoretical framework of brain function.

Why is studying action understanding and its neural bases important? We are social animals who interact with other people every day. Understanding others’ actions is fundamental for appropriate social behaviour. Recognizing others’ actions is considered a necessary ability for extrapolating their mental states (Frith & Frith, 1999) and predicting their intentions (Blakemore & Frith, 2005). Action understanding is also important for coordinating our own actions with those of other people (Kokal et al., 2009). It becomes obvious that action understanding has important consequences for everyday life when this ability does not function any more, either because it was impaired e.g. due to a stroke, or because it never developed properly. For example, it has even been proposed that action understanding impairment is at the basis of mental diseases such as autism (but see Hamilton, 2009; Williams et al., 2001). Investigating the neural bases of action understanding can provide us with profound knowledge about such diseases and the appropriate treatments.

The second reason for the widespread interest in this topic was the discovery of the so-called *mirror neurons* in the early 1990s by Rizzolatti’s group in Parma (Italy). These cells were observed in the monkey brain and have the property of discharging both during action execution and during the observation of this same action being performed by others (di Pellegrino et al., 1992). Because of this interesting property, many researchers have suggested that mirror neurons are involved in action understanding. According to this view, we are able to understand others’ actions because we activate those neurons that are active when we perform the same action ourselves. Since we know what we do when we perform the action, we also know what other people intend to do when we observe them performing the same action (Rizzolatti et al., 2001). Despite considerable support for this appealing theory, not everyone in the scientific community agrees with this view. An active debate on the role of mirror neurons in action understanding has emerged over the last few years (Caramazza et al., 2014; Csibra, 2007; Hickok, 2009; Mahon & Caramazza, 2008).

"What comes first? The understanding or the simulation?"

In the next sections of this chapter, I am going to introduce the two main classes of theories, which I will refer to as *cognitive* and *motor* theories (**Figure 1.1**), which attempt to explain the mechanisms underlying action understanding. Then, I will describe the different levels of action representations and why studying these different levels can help in investigating action understanding. Finally, I am going to describe why it is important to investigate the temporal aspects of action observation, and how this issue has been addressed in previous studies.

Theoretical background: models for action recognition

Behavioural studies have shown that motor knowledge is automatically retrieved during action observation. The kinematics of an action, namely the trajectories of vertical/horizontal arm movements, are affected if a performer observes an actor doing similar but incongruent actions (Kilner, Paulignan, & Blakemore, 2003). Evidence of motor knowledge retrieval has also been observed during sentence reading, demonstrating that motor representations are also activated when action-related concepts are triggered through language. Glenberg and Kaschak (2002) designed an experiment in which participants had to judge if action-related sentences were sensible or nonsense. For example, the sentence *"close the drawer"* is a sensible sentence and implies a movement away from the body. In half of the experimental blocks, participants could indicate that a sentence was sensible by pressing a button that was distal from the body (congruent movement); in the remaining blocks, the "sensible" button was more proximal to the body (incongruent movement). The reverse button-response mapping was assigned for sentences such as *"open the drawer"*. Participants' response time was faster if the response movement was congruent with the implied action in the sentence. Glenberg and Kaschak (2002) called this phenomenon the *action-sentence compatibility effect* (ACE) and argued that motor representations are automatically recruited in these kinds of tasks.

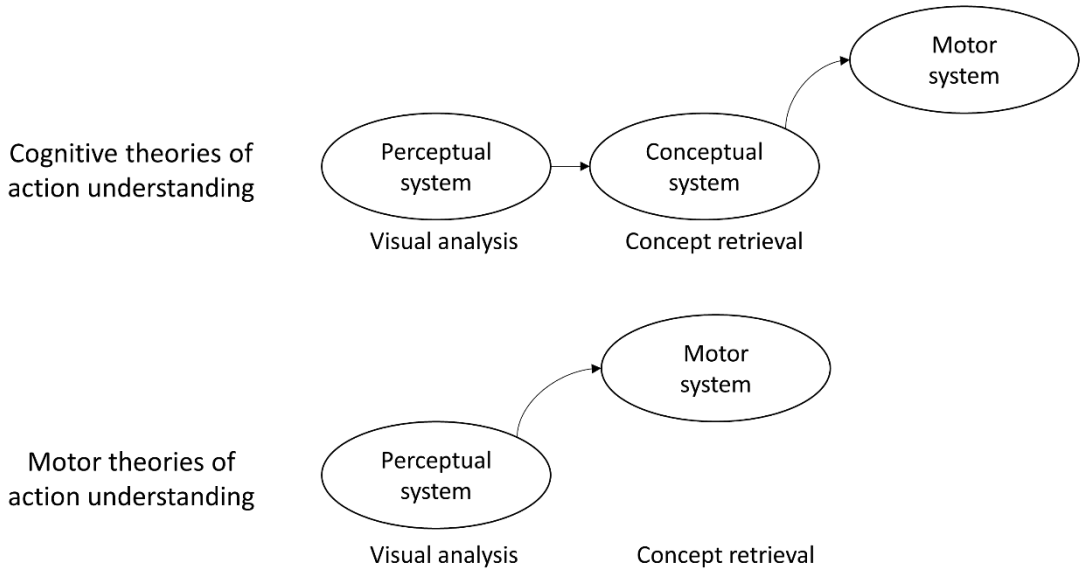


Figure 1.1. Models of action understanding. Schematic representation of the two major theories of action recognition. Cognitive theories (e.g. Mahon & Caramazza, 2008) state that action concepts are represented in a semantic system distinguishable from the motor system. Conversely, motor theories *(Rizzolatti et al., 2001)* claim that action semantics are directly encoded in the motor system. The two types of theories predict different neural dynamics of action understanding. Motor theory hypothesize a direct access to the motor system and consequently an early access to action semantics *(Hauk et al., 2008)*. By contrast, cognitive theories predict a post-conceptual encoding of action semantics in the motor system.

The majority of the experimental contributions showing motor-related activity during action observation come from studies that used neurophysiological and neuroimaging techniques such as direct single-unit recording in monkeys and functional magnetic resonance imaging (fMRI), electroencephalography (EEG), magnetoencephalography (MEG), and transcranial magnetic stimulation (TMS) in humans. Motor-related regions were found to be active during action observation, both in monkeys with direct neuronal activity measurement (di Pellegrino et al., 1992; Gallese et al., 1996) and humans using TMS (Fadiga et al., 1995). Interestingly, other areas are also recruited during action observation, including the parietal (in monkeys, Fogassi et al., 2005; and in humans, Shmuelof & Zohary, 2007) and temporal regions (in monkeys, Jellema,

Baker, Wicker, & Perrett, 2000; Perrett et al., 1989; in humans Decety et al., 1997; Lingnau & Petris, 2013; Oosterhof, Tipper, & Downing, 2012a; Oosterhof, Wiggett, Diedrichsen, Tipper, & Downing, 2010; Giacomo Rizzolatti, Fadiga, & Gallese, 1996; Wurm, Ariani, Greenlee, & Lingnau, 2015; Wurm & Lingnau, 2015; see also Caspers, Zilles, Laird, & Eickhoff, 2010). More specifically, these regions include the ventral premotor (PMv) cortex, the inferior parietal lobule (IPL) and the posterior middle temporal gyrus, and have been identified as the *action observation network*, or AON (see Caspers et al., 2010 for a detailed meta-analysis of fMRI studies). **Figure 1.2** shows the regions typically reported in action observation studies.

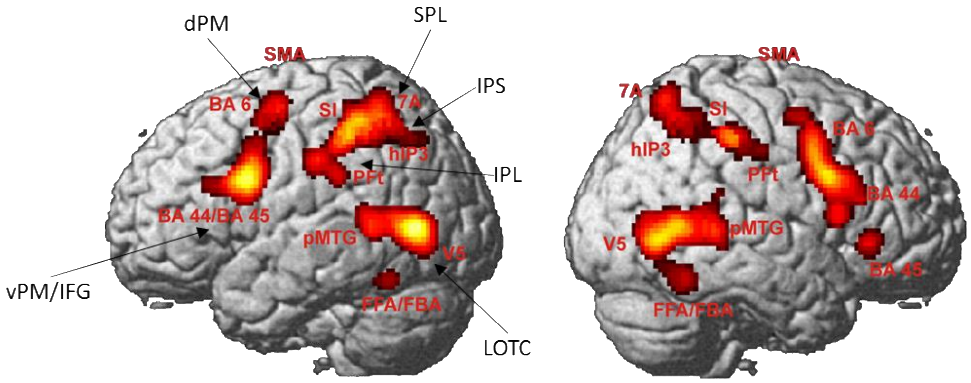


Figure 1.2. The action observation network (AON). Regions typically recruited during action observation tasks. The network includes temporal, parietal and frontal areas. The image shows the results of a meta-analysis of 139 fMRI studies (Caspers et al., 2010). **BA44/BA45**, inferior frontal gyrus; **BA6**, dorsal premotor cortex; **SMA**, supplementary motor area; SI, primary somatosensory cortex; **7A**, superior parietal area; **PFt**, inferior parietal area; **hIP3**, intraparietal area; **pMTG**, posterior middle temporal gyrus; **V5**, visual area 5; **FFA/FBA**, fusiform face area/fusiform body area. Left panel, added in black, alternative labels used in this thesis: **vPM**, ventral premotor; **IFG**, inferior frontal gyrus; **dPM**, dorsal premotor; **SPL**, superior parietal lobule; **IPS**, intraparietal sulcus; **LOTC**, lateral occipitotemporal cortex. Figure modified from (Caspers et al., 2010).

Both motor and cognitive theories are compatible with the behavioural results previously described and they both predict the recruitment of motor and nonmotor regions (Mahon & Caramazza, 2008; Rizzolatti et al., 2001). The debate arises when assigning the central role in action understanding to one or another of these regions. The motor theory argues that the motor system plays a crucial role in action understanding, whereas the cognitive theory claims that the meaning of an action is

processed outside the motor system and the involvement of motor regions is a consequence of action understanding (see **Figure 1.1**).

Motor theories of action understanding

The motor theory is part of a larger group of *embodied* theories of cognition, also known as *grounded cognition theories*, which claim that higher cognitive abilities are the result of the reactivation of those areas that are directly dedicated to primary sensory processing or to action execution (Barsalou, 2008; for a critical assessment see Caramazza et al., 2014). At the neural level, higher-level cognition is not due to computation within amodal or abstract brain regions that are independent of the perceptual and motor system. Cognition is *grounded* in the low-level perceptual and motor systems. One classical example of embodied theory is the *motor theory of speech perception* (Liberman et al., 1967), which claims that speech is perceived and understood through internal motor simulation of those *articulatory* gestures used during the production of the same speech sounds.

Similarly, motor theories claim that we understand others' actions by means of the re-enactment of those neural substrates and simulation of the same processes that are active when we perform the same actions (Pulvermüller, 2013; Rizzolatti & Sinigaglia, 2010; Rizzolatti et al., 2001). For example, when we observe a person grasping a cup, we activate the same brain regions (hand-related motor and premotor areas) that are active when we grasp a cup; when we see someone kicking a ball, we activate the same brain regions (foot-related motor and premotor areas) that are active when we kick a ball. These *shared* neural representations between observed and executed actions are considered to contain the *knowledge* (we know what we do when we are grasping or kicking) that allows us to understand the observed actions (Pulvermüller, 2005; Rizzolatti et al., 2001). Action semantics (or action knowledge in general) are grounded in the motor system.

This logical and simple explanation received strong support from the discovery of mirror neurons (di Pellegrino et al., 1992). In fact, these theories are sometimes referred to as the *mirror neuron theory of action understanding*. The core assumption

of motor theories is that there is no need for an intermediate cognitive step between action perception and action understanding: once the observed action is mapped onto our own motor representation, we grasp the meaning of this action. It has been argued that this mapping occurs quite quickly, within 200 milliseconds (Boulenger et al., 2006; Hauk & Pulvermüller, 2004), which has been taken as a support for the idea that motor-related regions are necessary for action comprehension.

Similar evidences are provided by the language domain. According to the embodied account, action words are strictly linked to the motor system, because words refers to the action semantics located in motor-related areas (Pulvermüller, 2005). Therefore, when we read or listen to effector-specific action words, effector-specific brain areas get involved and provide us access to the word semantics. For example, reading the word ‘*kick’* activates foot motor regions and we can thus understand the meaning of the word.

From an experimental point of view, the five main predictions of the motor theory of action understanding are: 1) action observation leads to activation of motor-related regions, specifically those areas involved when the observer performs the same action; 2) the action representation encoded in motor-related areas generalizes across different action implementations; 3) the motor system has an early access to abstract action representations; 4) impairment of motor-related regions affects action recognition; 5) impairment of non-motor-related regions does not affect action recognition.

In this thesis, I will not cover the predictions related with impairment of motor and nonmotor regions (point 3 and 4). Patient studies have provided mixed results on this matters, but recent studies appear to consider motor-related regions not necessary for action understanding (Negri et al., 2007; Tarhan et al., 2015). Detailed information are provided elsewhere (e.g. Buxbaum & Kalénine, 2010; Hickok, 2009; Kalénine et al., 2010; Negri et al., 2007; Pazzaglia et al., 2008; Tarhan et al., 2015).

Cognitive theories of action understanding

An alternative explanation is that activation of motor-related areas during action observation is *a consequence* of action understanding that occurs in so-called *abstract* nonmotor regions, as stated by the *cognitive theory of action perception*. The basic idea is that such regions would contain action-related information (action knowledge) that is not linked to specific motor programmes (Caramazza et al., 2014; Hickok, 2012; Mahon & Caramazza, 2008). These regions form a sort of *conceptual system*, distinguishable from the perceptual and motor systems, from which action information is retrieved any time an action concept is needed. Cognition occurs through computation within amodal and abstract regions. However, the semantic system is highly interconnected with the perceptual and motor systems such that activating a specific concept might lead to re-activation of the interconnected systems. For example, the concept *grasping* would be active during action execution, action observation of grasping actions and reading of action-related word. In all cases, we *understand* the action we are performing or observing and we activate the same concept. The difference concerns *how* the concept is recruited: through internal processing in the case of action execution and by visual external stimuli during action observation. The question is if re-enactment of motor-related regions has a functional role. In the last several years, several different, but not mutually exclusive, explanations have been suggested. Heyes et al. (2010) and Catmur et al.(2007) proposed an association learning account: action execution and perception mostly co-occur in everyday life and this automatically creates sensory-motor associations. This account could explain why motor system activity correlates with motor experience (Calvo-Merino et al., 2005, 2006; Cross et al., 2006; for a review see Turella et al., 2013). Furthermore, it has been suggested that mirror neurons might be helpful for selecting appropriate actions (Hickok & Hauser, 2010), similar to *canonical* neurons which discharge for objects according to their manipulation properties (Rizzolatti et al., 1988). Finally, mirror neurons might be part of a two-pathway (dorsal and ventral) system and their function could be to generate the sensory consequences of concrete actions that are understood within temporal regions where abstract action

representations are encoded (Kilner, 2011). In any case, the cognitive theory makes similar predictions to the motor theory: action observation leads to activation of motor-related regions.

However, the cognitive theory predicts no impairments in action recognition in those patients that have deficits in action production. Some experimental evidence is in line with this prediction. The most famous example is Broca’s aphasia. People with damage in Broca’s area (within the inferior frontal gyrus), which is directly related to speech production, can still understand action words, and words in general, even if they are unable to reproduce those words (Broca, 1861). In the action observation domain, a recent patient study by Buxbaum’s group (Tarhan et al., 2015) using voxel-based lesion symptom mapping on 131 chronic left-hemisphere stroke patients found that impairment of action recognition is related to damage within lateral-occipitotemporal areas, whereas damages to primary motor and somatosensory cortices and inferior parietal lobule accounted for action execution deficit. Furthermore, an anterior section of LOTC was related to impairment in both action execution and recognition. These results suggest that actions can be represented at different levels of abstractions in the brain.

Levels of action representations

Actions can be described at different levels (**Figure 1.3**). If we reconsider the initial example about making coffee, we can identify a low-level description of the action (e.g. the actual movements/steps required to make coffee using the moka pot) and move along a continuum towards an abstract description (making coffee). We can even distinguish proximal actions (filling the moka-pot) from distal ones (drinking the coffee). It is likely that these different levels of representations are also reflected at the neural level (Grafton & Hamilton, 2007; Hamilton & Grafton, 2008; Kilner, 2011).

It is important to note that the activation of the same motor-related regions during action observation is a necessary *but not sufficient* condition for proving that action knowledge resides within the motor system. In fact, fMRI studies have reported many

regions, also outside the classical action observation network, that seem to show a shared common code during action observation and execution (Gazzola & Keysers, 2009). However, is this activation really reflecting action comprehension? Other cognitive processes, such as attention, response selection or task engagement can explain this functional neural overlap (Dinstein et al., 2008).

A region that is involved in action understanding should encode abstract representations of actions (Hamilton & Grafton, 2008; Hickok, 2012; Oosterhof et al., 2013). In other words, a region dealing with action meaning does not need to discriminate between the ways in which an action is performed (e.g. the kinematics). For example, *grasping* can be performed in many different ways (e.g. with the right or the left hand) and can be observed from a potentially infinite number of perspectives, but the goal always remains the same. Which region contains information concerning the meaning of actions? Neuroimaging studies have shown that abstract actions are represented in temporal and parietal regions (Oosterhof et al., 2010, 2012a; Wurm & Lingnau, 2015; Wurm et al., 2015), but also in frontal areas, specifically in premotor regions (Cattaneo et al., 2010).

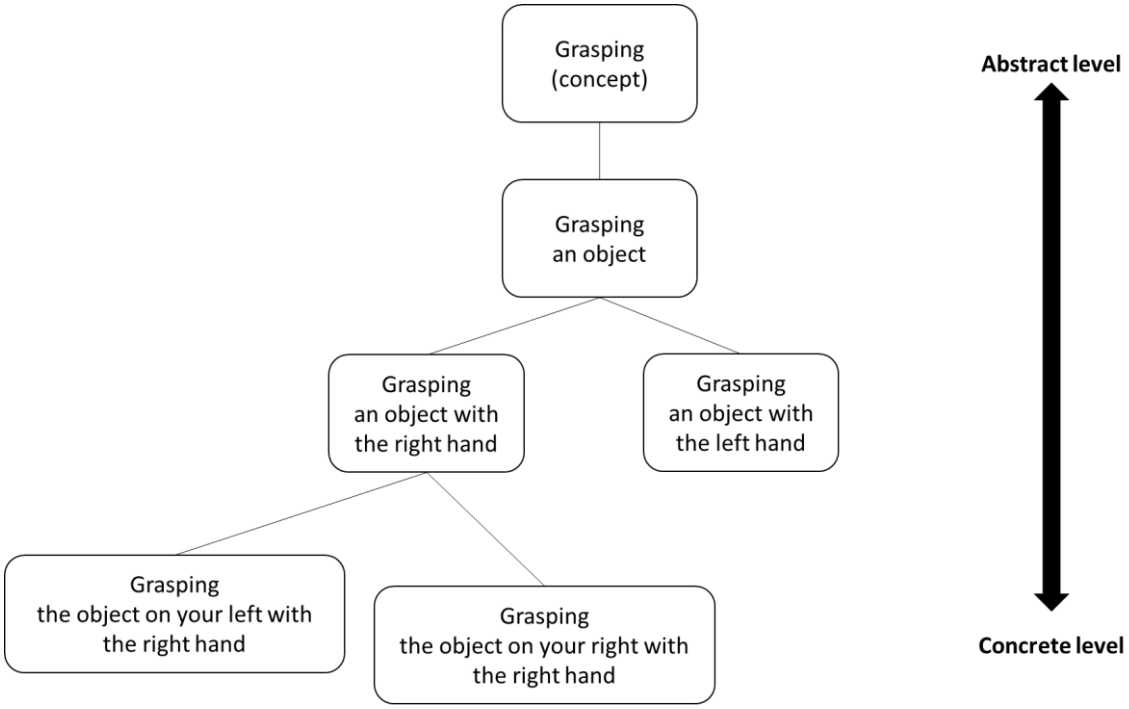


Figure 1.3. Hierarchical organization of actions. Action representations considered as a concrete to abstract hierarchy of representations. The lowest level describes the concrete aspects of an action (e.g. the effector used, the position of the object, the perspective of the scene, etc.). Increasingly abstract representations arise at intermediate levels of the hierarchy while the highest level represents the action concept.

Looking for brain regions that encode abstract information on actions can help to reveal the role of the areas typically involved in action observation. For example, some regions might be encoding specific low-level features of observed actions (e.g. the effector used, or movement direction). Other regions might ignore this type of information and instead encode the goal associated with them. Recent methods of data analysis such as *multivariate pattern analysis* (MVPA) applied to fMRI can contribute to this type of investigation by extrapolating the informational content of a brain region (Haxby et al., 2001; Kriegeskorte et al., 2006; see Haxby, 2012 for a historical perspective). Recently, multivariate analyses have also been applied to MEG data (King & Dehaene, 2014). In the studies I am going to describe in this thesis, I adopted both MVPA of MEG data (**Study 1** and **Study 2**) to investigate abstract action

representations over time and MVPA of fMRI data (**Study 3**) to characterize what is encoded within regions involved in action observation.

Temporal aspect of action observation

Action understanding occurs in at least two separate and sequential stages (**Figure 1.1**): 1) visual analysis of the action information, and 2) action semantics retrieval. According to the motor theory, the second step coincides with motor information retrieval. According to the cognitive account, motor system activity is a third step subsequent to the retrieval of action semantics. The temporal aspect of action understanding, together with the hierarchical organization of action information in the nervous system, provides a powerful tool for testing the two theories. If abstract action representations are directly mapped to the motor system of the observer, it would be hard to reconcile with cognitive theories of action understanding. On the other hand, if we observe a two-step process (action semantics retrieval followed by motor representations), then we have to reject the motor account. In the literature, several studies have investigated neural dynamics during action observation, but to date none has focused on the dynamics of abstract action representations.

Ritta Hari’s group conducted the first studies that aimed at investigating the temporal dynamics of action observation. They used MEG to measure the magnetic activity generated by neural sources involved during action observation and to track their temporal recruitment. They observed different event-related field (ERF) components emerging at different latencies for different sources. The earliest ERF peak started in left lateral occipital areas (around 250 milliseconds before the hand touched the object) followed by premotor ventral regions (around 100 millisecond before the hand touched the object) and finally to primary motor regions (just after the hand touched the object). This kind of temporal pattern was observed both during hand-related (Nishitani & Hari, 2000) and mouth-related (Nishitani & Hari, 2002) actions. These studies confirmed that motor regions are recruited during action observation, adding important information about the temporal dynamics of such recruitment. Moreover,

these studies reveal a temporal profile that is similar to the one described in the monkey brain during action observation (Fabbri-Destro & Rizzolatti, 2008). However, they do not provide much information about *when* action understanding occurs or *which* region might be responsible for action recognition. To do so, one needs to test which of the involved region encodes abstract action information (see previous section *Levels of action representations*).

In addition to event-related components, EEG and MEG allow us to measure brain oscillatory activity. Neuronal activity generates different spontaneous rhythms that can be measured at the scalp. These rhythms can be classified according to their frequency, topography and modulation due to external stimuli or internal brain states. The frequency bands that are typically modulated during action observation range between 8-12 Hz (alpha band) and between 18-25 Hz (beta band) and are usually observed at sensors located over sensorimotor regions (they are also called mu rhythms; see Pineda, 2005 for a review). The perturbation due to the presentation of action-related stimuli is usually referred to as mu-band activity. It is considered an induced response that has been suggested to reflect “the changes in activity of the local interactions between main neurons and interneurons” (Pfurtscheller & Lopes da Silva, 1999). In studies investigating oscillatory brain activity during action observation, the main dimension of interest is the power of the frequencies elicited at specific time points during measurement. The idea is that mu rhythms are the neural correlates of the motor system at rest (Pfurtscheller, Stancàk, & Neuper, 1996; but see Engel & Fries, 2010) and the decrease in power (event-related desynchronization or ERD) observed during action observation would indicate motor system activation, whereas power increase would be the neural marker for motor system inhibition (event-related synchronization or ERS) (Hari, 2006; Pineda, 2005).

It has been shown, using EEG, that observation of a grip action causes mu desynchronization (Muthukumaraswamy & Johnson, 2004) and that this desynchronization is stronger when the action targets an object (Muthukumaraswamy et al., 2004). These studies corroborate the idea that the motor system is active during

action observation and suggest that brain oscillatory activity is a suitable tool to study action observation. However, the authors focused only on the frequency dimension of the mu-band activity without considering the temporal aspect of the suppression. When the motor system is involved in the process of action observation and how the other regions of the AON are recruited is still a matter of debate. Three recent studies have started directly investigating the temporal dynamics of action observation in more detail. In 2008, Koelewijn and colleagues (2008) investigated the oscillatory activity correlated with the observation of an apparent movement of the hands and reported that sources located in premotor regions were modulated by the incorrectness of the observed action. Pavlidou et al. (2014a) explored the temporal dynamics associated with the observation of actions simulated by point-light displays (PLD). PLD stimuli are generated by recording biological movements using motion-capture devices positioned on the major joints of the body of an actor. In this way, it is possible to create human-like figures performing actions where only the essential features necessary to perceive the biological movement are maintained (Johansson, 1973). By contrasting action stimuli with scrambled PLDs, Pavlidou and colleagues found alpha- and beta-activity over premotor regions, as expected, and over parieto-occipital and temporal regions. They were able to describe in more detail the process of action observation as actions unfolded, exploring the neural correlates that are likely associated with the activity of the AON. Source analysis revealed many regions typically considered to form part of the AON aside from low-level sensory regions like V1 and M1. Interestingly, temporal regions, such as the superior temporal sulcus and middle temporal areas, were modulated earlier than frontal regions. However, their results show that these regions are somehow involved in action observation, but do not say which region encodes an abstract representation of the action.

Multivariate analysis in combination with MEG can be a powerful tool for studying abstract action representations. Oscillatory activity might contain information related to different aspects of an action (e.g. the goal) and these different representations can be decoded using MVPA. This is exactly what I did in my work (Study 1 and Study 2).

“What comes first? The understanding or the simulation?”

Current work

During my PhD, I have investigated the neural dynamics of abstract action representations during action observation with a particular focus on temporal regions with the aim of revealing the action-related properties encoded within these regions. In this thesis, I am going to present three studies (two MEG and one fMRI) that I have conducted in these four years. These three experiments aimed at answering three main questions: 1) When, and in which brain regions, are abstract action (grasping/pointing) representations encoded (**Study 1**, in Chapter 2)? Do the results observed in Experiment 1 generalize across other actions (**Study 2**, in Chapter 3)? What aspect of an action is encoded within temporal regions (**Study 3**, in Chapter 4)?

To anticipate my findings, the MEG experiments revealed that lateral occipitotemporal cortex (LOTC) has the earliest access to abstract action representations (**Study 1**,) and that this result is generalizable to different types of actions (**Study 2**). Furthermore, LOTC seems to encode different aspects of actions (such as the state of a manipulable object), making this region a good candidate for hosting the semantics of actions (**Study 3**).

Chapter 2: MEG-MVPA demonstrates that LOTC encode abstract action representations earlier than frontal regions

Adapted from:

Tucciarelli R., Turella L.,Oosterhof N.N., Weisz N., and Lingnau A.(2015). *MEG multivariate analysis reveals early abstract action representations in lateral occipitotemporal cortex.* Journal of Neuroscience.

Introduction

How do we assign meaning to actions performed by other people? One of the most dominant views in the literature is the idea that action concepts are grounded in the motor system (Kiefer & Pulvermüller, 2012; Rizzolatti et al., 2001). By contrast, according to classical cognitive theories (Caramazza et al., 2014; Mahon & Caramazza, 2008), the ability to understand the meaning of other people's actions draws on conceptual representations stored outside the motor system, such as posterior temporal regions.

A region involved in action understanding should be able (a) to discriminate between different actions (action specificity), and (b) to generalize across different possible instances of a particular action (Caramazza et al., 2014; Dinstein et al., 2008; Oosterhof et al., 2013). For example, grasping has the same meaning for an observer irrespective of whether the movement is performed with the left or right hand, or towards the left or right side of visual space. In other words, a region important for action understanding should represent the action while generalizing across concrete

instantiations such as the underlying effector or reach direction. Previous fMRI and TMS studies in humans have reported abstract action representations in parietal, frontal and occipital regions (Cattaneo et al., 2010; Hamilton & Grafton, 2006, 2008; Oosterhof et al., 2010, 2013), making it difficult to draw firm conclusions regarding the ongoing debate between motor and cognitive theories. The underlying temporal profile of action understanding is an important factor which is not yet fully understood. Such information is crucial since the two theories lead to opposite predictions: according to motor theories, motor areas should have the earliest access to abstract action representations (Pulvermüller, 2005). By contrast, according to cognitive theories, areas outside the motor system should have the earliest access to such abstract action representations.

Here we use MVPA of MEG data to examine where in the brain and at which point in time it is possible to distinguish between observed pointing and grasping irrespective of reach direction (left, right) or effector (left, right hand). In contrast to motor theories of action understanding, we show that abstract action representations are encoded in lateral occipitotemporal cortex earlier than in precentral regions.

Materials and Methods

We carried out two separate experiments with two different groups of participants: one behavioural experiment to identify the time point at which the videos contained enough information to allow participants to discriminate between pointing and grasping, and an MEG experiment. The same stimuli were used for the two experiments.

Participants. Fourteen students (7 females; mean age: 23.13; sd: 2.253; all right-handed) from the University of Trento took part in the behavioural experiment and received a reimbursement of €6, at the end of the session. A different group of seventeen students (11 females; mean age: 23.3; sd: 2.1; all right-handed) from the University of Trento with normal or corrected-to-normal visual acuity and with no neurological disorders took part in the MEG experiment. All participants received a

reimbursement of €25,- at the end of the MEG session. All of them gave informed consent in accordance with the Declaration of Helsinki. The experimental procedures were approved by the Ethics Committee for research involving human participants at the University of Trento.

Stimuli. Stimuli consisted of short video-clips (833 ms) depicting simple center-out hand movements (**Figure 3.1A**). Each clip started with the hand of the actor touching the central object (a polystyrene semi-sphere) with the index finger resting in the same position. After a variable amount of time (median: 183 ms; range: [67 – 367 ms]), a center-out movement towards one of the other semi-spheres started. Movement onset was defined as the time point in which the rest position was released and hand preshaping was initiated. The video ended as soon as the hand reached one of the peripheral semi-spheres (see **Figure 3.1A** for an example trial sequence). The actions were recorded from four different actors (1 male) using a digital video camera. Only the hands (and part of the forearm) of the actors were visible in the field of view. We instructed the actors to keep the velocity and kinematics of the movements as similar as possible across the two different movements. We discarded, based on our perceptual judgment, videos in which the velocity or kinematics were too dissimilar from the others and videos in which the preshaping of the hands before movement onset could give information regarding the upcoming action, keeping a total of 80 videos (five exemplars for each combination of actor x movement type x direction). We obtained movements performed with the left hand creating specular copies of the right hand movement videos via software (Matlab, Mathworks, Natick, NA), for a total of 160 videos. On each video, we superimposed a small white cross (0.88 x 0.88°) above the central semi-sphere to enable fixation and thus to avoid possible noise in the MEG signal due to eye movements.

MEG-MVPA demonstrates that LOTC encode abstract action representations earlier than frontal regions

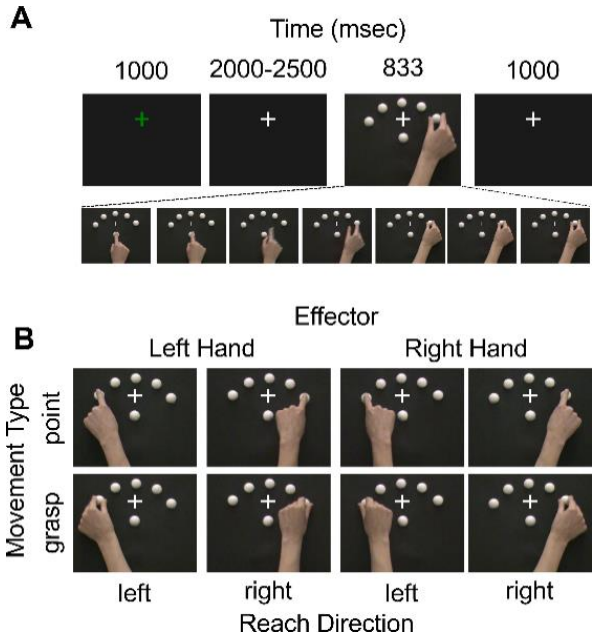


Figure 2.1. Example of a trial sequence and experimental design. During MEG recording, N = 17 participants watched video clips of simple 'reach-to-point' or 'reach-to-grasp' movements (duration: 833 ms). Participants were instructed to fixate on a central fixation cross while attentively observing the entire video without performing any movements. To ensure that participants paid attention to the videos, different types of questions were asked during occasional catch trials that were later discarded from the analysis (for details, see Material and Methods). The green fixation cross indicated the period during which participants were allowed to blink. Eye movements were recorded using an MEG-compatible eye-tracker. B: We used a 2x2x2 design, manipulating the type of movement (pointing/ grasping), reach direction (left/ right), and effector (left/right hand).

Behavioural experiment

Procedure. To identify the minimum video duration required to be able to distinguish between observed pointing and grasping, we presented participants with videos depicting pointing or grasping movements directed towards the left or right side, performed with the left or right hand. The duration of the videos was parametrically varied (167, 200, 233, or 333 ms). Participants had to classify the type of observed movement by pressing one of two possible buttons while ignoring the other two dimensions (reach direction, effector). A trial started with a fixation period (white cross) of 2 seconds. Then the video appeared for a variable duration. As soon as the video ended, the fixation cross appeared again, and participants had to indicate by button press which movement they had observed. Participants were instructed to respond as accurately as possible. Video duration, type of movement, effector and

reach direction were randomized. Each participant completed four experimental runs of approx. 5.5 min, for a total of 512 trials (64 repetitions per conditions). Stimuli were presented on a CRT monitor (ViewSonic Graphic Series G90fB; screen resolution: 1280x1024, refresh rate: 60 Hz) placed approx. 64 cm in front of the participants.

Statistical analysis. The aim of the behavioural experiment was to individuate the point in time in which the two actions started to perceptually diverge. To compute the accuracy for discriminating between the two observed actions as a function of video duration, we divided the number of correct classifications by the total number of trials, separately for each video duration and each participant, collapsing across effector (left, right hand) and reach direction (left, right). We then used a chi-squared test to assess at which video duration the accuracy was higher than chance level (50%).

MEG experiment

Procedure. We presented participants (N = 17) with short videos (833 ms) of reach-to-point and reach-to-grasp movements performed with either the left or right hand towards peripheral targets on the left or right side (**Figure 3.1A**) while measuring their brain oscillatory activity. We used a 2x2x2 factorial design (**Figure 3.1B**), varying the type of movement (pointing/grasping), the effector (left/right hand), and reach direction (left/right). Each trial consisted of the following events (**Figure 3.1A**): a green fixation cross (blink phase: 800 ms), a white fixation cross (fixation phase: randomly jittered within 2000-2500 ms), the video (video phase: 833 ms), and a white fixation cross (resting phase: 1000 ms). Trial duration varied from 4633 ms to 5133 s depending on the duration of the fixation phase. The blink phase at the beginning of each trial allowed participants to blink during a controlled time window and thus reduced the probability of blinking during the fixation phase or during video presentation. Participants were instructed to blink every time they saw the green cross. During the fixation phase, participants had to maintain fixation on the white cross. We jittered the fixation phase to prevent participants from predicting the appearance of the video since this could cue a neural response. When the video appeared, participants were asked to keep fixating on the cross and to pay attention to the ongoing movement. In

particular, in contrast to the task used in the behavioural experiment, we asked them to attend to all three dimensions we manipulated, i.e. movement type, effector, and reach direction. During the resting phase, participants had to keep fixating and to wait for the green cross that indicated the beginning of a new trial.

To ensure that participants were paying attention to the video, we introduced catch trials (10% of all trials) during which we presented a question regarding one of the three dimensions (e.g. 'was the direction to the left?'). Catch trials were presented occasionally with the following constraints: 1) if trial N was a catch trial, trial N+1 could not be a catch trial; 2) no catch trial during the first trial of a run. A catch trial was identical to an experimental trial except for the question that appeared at the end of the catch trial (1 sec after video offset). Since participants did not know when a catch trial would appear, and what the question would be, they had to pay attention to each video and to each of the three dimensions to perform the task correctly. The answer was always binary (yes or no) and participants had MEG-compatible buttons for answering to the questions. The assignment of the response to the two buttons changed randomly for each question to avoid any potential confounds related to motor preparation. Eye movements were monitored using the OEM system (OEM eye tracker, SMI; 60 Hz sampling rate). After each response, feedback was provided (a smiling or a sad cartoon face).

Each participant performed 10 runs, consisting of 64 trials, plus 6 catch trials, for a total of 640 experimental trials and 60 catch trials. The number of repetitions for each factorial combination (movement type x effector x reach direction) per participant was 80. Before entering the shielded room, participants were familiarized with the stimuli and the task. Each run lasted from 4.9 to 5.5 minutes, depending on the duration of the fixation phase, for a total session duration of approximately 52 minutes. At the end of each run, participants were allowed to rest for a few minutes before a new acquisition started.

Stimuli were projected on a screen (1280 x 1024 pixels screen resolution, 60 Hz refresh rate) that was placed about 130 cm in front of the participant. The screen was visible

as a rectangular aperture of about 21.7 x 13.16°. We controlled visual stimulation during the behavioural and the MEG sessions using ASF (Schwarzbach, 2011), a toolbox for Matlab (Mathworks, Natick, MA) based on the Psychtoolbox (Brainard, 1997).

MEG data acquisition and analysis. At the beginning of the MEG session, the head shape of each participant was digitally acquired using the Polhemus system (Polhemus, Colchester, VT). Moreover, we placed three coils at the participant's forehead and two behind the ears to acquire the head position of each participant within the MEG helmet at the beginning of each run. Prior to entering the shielded room containing the MEG system, participants were asked to remove all magnetic materials that could distort the measurement.

We measured neuromagnetic brain activity using a 306-channels whole head MEG system (Neuromag Elekta Oy, Helsinki Finland) at a sampling rate of 1000 Hz. The system consists of 204 planar gradiometers and 102 magnetometers arranged in a helmet configuration. Here, we only report the results of the gradiometers. Triggers were sent at video onset to synchronize stimulus presentation with neural activity. To check for the correct timing of the stimuli, and to take into account possible delays of the stimulus presentation with respect to the triggers, we used a photodiode on the stimulation screen inside the shielded room.

MEG data preprocessing. We analysed data using the open source Matlab-based Fieldtrip toolbox (Oostenveld et al., 2011). Continuous data were cut into epochs from -1 to 1.3 sec relative to video onset. Epochs were high-pass filtered at 1 Hz to remove very slow frequencies and Direct Current (DC) offset. Frequencies due to the electrical system were also filtered out using a band-stop filter (Butterworth IIR filter) at 50 Hz and its harmonics (100 and 150 Hz). Trials with blinks or eye movements during the presentation of the video or during the baseline period were discarded on the basis of the information from the eye tracker. In addition, we visually inspected trials for artifacts, blind to the condition, and rejected trials that were clearly affected by external noise or spike current. On average, we rejected 13% of the trials per participant. If a sensor was very noisy for the entire experimental session, it was

rejected. In order to have the same number of sensors for each participant, missing sensors were reconstructed by interpolation from neighbours.

Time-frequency analysis. To obtain a time-frequency representation of the oscillatory activity associated with movement observation, we applied Fourier transformation to sliding time windows with fixed length of 500 milliseconds. The sliding window moved in steps of 50 ms; power was calculated for frequencies in a range from 2 to 40 Hz in steps of 2 Hz. To avoid spectral leakage and to control for frequency smoothing, a Hanning taper was applied prior to Fourier transformation. Subsequently, for the univariate analysis only, power was averaged across effector and reach direction, and the spectral power was normalized relative to baseline (-0.5 to -0.3 s with respect to the onset of the video, i.e. during a sub-period of the fixation phase).

Source analysis. Neural sources were found using dynamic imaging of coherent sources (DICS), a frequency domain beamforming technique (Gross et al., 2001). We chose the frequencies and times of interest based on the sensor level analysis. Specifically, we considered the sensor with the greatest accuracy of the classifier (multivariate analysis) to distinguish between pointing and grasping, generalizing across effector and reach direction, in those frequency bands that survived the multiple comparison tests. Note that, given the way the sensors were selected, source analysis merely served as a visualization of the sources.

For each participant we used a volume conductor model using the single-shell method (Nolte, 2003). The models were built warping a dipole grid based on a MNI template brain to fit the individual head shape of each participant. We proceeded with DICS for each separate condition using a common spatial filter computed from the combination of the two conditions. In this way, any difference between the two conditions cannot be ascribed to differences between the filters.

MEG Statistical analysis (sensor level). We carried out both uni- and multivariate analyses in sensor space, followed by a beamforming analysis (Gross et al., 2001) to identify sources explaining any observed effects. Univariate analysis was conducted in order to observe the classical decrease in power in alpha and beta bands (Cochin et al.,

1999; Hari, 2006; Pineda, 2005). Importantly, to identify at which sensors and at which point in time it is possible to distinguish between the two movements on the basis of the MEG signal, we applied multivariate analysis on the computed power and the sources adopting an algorithm developed for the analysis of fMRI data (Oosterhof et al., 2012a).

Behavioral analysis (MEG experiment). Participants' accuracy in answering the questions in the catch trials during the MEG experiment was evaluated online by observing the feedback provided after each catch trial. All participants were able to answer the questions and typically made two or three mistakes within the entire session (mostly at the beginning of the experiment). We are thus certain that participants were attending to the videos.

Univariate analysis. Note that in contrast to the multivariate analysis, in which we specifically targeted regions that show movement selectivity generalizing across effector and reach direction, the purpose of the univariate analysis was to identify areas with less specific properties. In particular, as a quality control, we examined whether we obtained the typical decrease in the alpha and beta band during action observation (Cochin et al., 1999; Hari, 2006; Pineda, 2005). Furthermore, we aimed to determine which frequency bands and which sensors were modulated differently during pointing and grasping when collapsing across effector and reach direction.

All the experimental conditions were baseline corrected by subtracting the fixation period (from -0.5 to -0.3) from the post-stimulus period (from 0 to 1.3 s). To assess the difference between pointing and grasping, we used a non-parametric method (permutation test), with a cluster method for multiple comparison correction (Maris & Oostenveld, 2007) with participants as units of observation. In brief, we computed t scores between the two movements for each sensor-frequency-time bin. The observed cluster-level statistic was obtained by summing the t scores of neighboring bins (in time, frequency and sensors) exceeding an a priori defined critical value ($p<0.05$). We repeated the procedure 1000 times by swapping the condition labels and we obtained the distribution of permuted cluster-level statistics. At each iteration, the maximum

cluster-level statistic was considered to control for type I error. The p value was the proportion of permuted cluster-level statistics that exceeded the observed cluster-level statistic. If the p-value was less than 0.05, the cluster was taken as significant.

Multivariate analysis. The assumption behind multivariate analysis in MEG is that the processing of each stimulus category is associated with a specific neural activity that induces an oscillatory signal (or neural pattern) consisting of a unique combination of sensor, time, and/ or frequency. Multivariate analyses exploit differences in terms of these patterns of activations. By contrast, univariate analyses do not consider such patterns, but address whether two conditions differ in terms of the average response of a single variable (e.g. averaged frequency over time). This is why multivariate analyses are more sensitive than univariate analyses (Haxby et al., 2001; Kriegeskorte et al., 2006). Importantly, multivariate analysis allows analyzing whether the representational content of an area – examined via the underlying neural pattern - generalizes across low-level features. In our case, we aimed to identify regions in which the unique neural patterns associated with pointing and grasping generalized across effector (left or right hand) and reach direction (left or right; for a schematic overview, see **Figure 3.2**). We trained a classifier to discriminate between the two types of movements using the spatio-spectral-temporal MEG signal (for details, see next paragraph) related to movements performed with one of the two effectors and towards one of the two directions. We then tested on the opposite combination of effector and direction. For example, we trained a classifier to distinguish between observed grasping and pointing actions performed with the left hand towards the left, and tested the classifier to distinguish between observed grasping and pointing performed with the right hand towards the right. In this way, above chance classification could only be due to information related to the type of movement, and not to low-level perceptual features.

Analyses were performed using CoSMoMVPA, an MVPA toolbox in Matlab (Oosterhof et al., in preparation). The toolbox provides an adapted version of the multivariate searchlight approach (Kriegeskorte et al., 2006), an information-based algorithm that

allows analyzing the neural contents adopting a multivariate approach at each location in the brain. In this analysis, we used local 'neighbourhoods' of features in channel-time-frequency space. We used a sensor radius of 1, a time radius of 100 ms, and a frequency radius of 8Hz. For a given 'centre' feature [a (sensor, time, frequency)-triple], its neighbours consisted of features for which its sensor, time, and frequency where all within the corresponding radii.

The main steps used in the multivariate analysis (for a schematic illustration, see **Figure 3.2**) were: 1) compute the time-frequency representation separately for each sensor and each trial (**Figure 3.2A**); 2) select the 'central' feature and its neighbours in time-frequency-sensor space (dashed rectangles in the insets in **Figure 3.2A**; for an enlarged view, see **Figure 3.2B**); 3) create a feature vector for each trial by selecting all features in its neighbourhood (**Figure 3.2C**) and normalize (z transform) the data; 4) create independent partitions for training and testing the classifier (see **Table 2.1**); 5) train the classifier; 6) test the classifier. We repeated the steps from 2 to 5 for each sensor and for each time and frequency bin, and the classification result for each central feature was assigned to its corresponding location in time-frequency-sensor space. For classification, we used a Support Vector Machine (SVM) algorithm, a type of classifier that looks for linear combinations of features to create a decision boundary to discriminate between two classes or stimuli (e.g. Mur et al., 2009; Pereira et al., 2009).

MEG-MVPA demonstrates that LOTC encode abstract action representations earlier than frontal regions

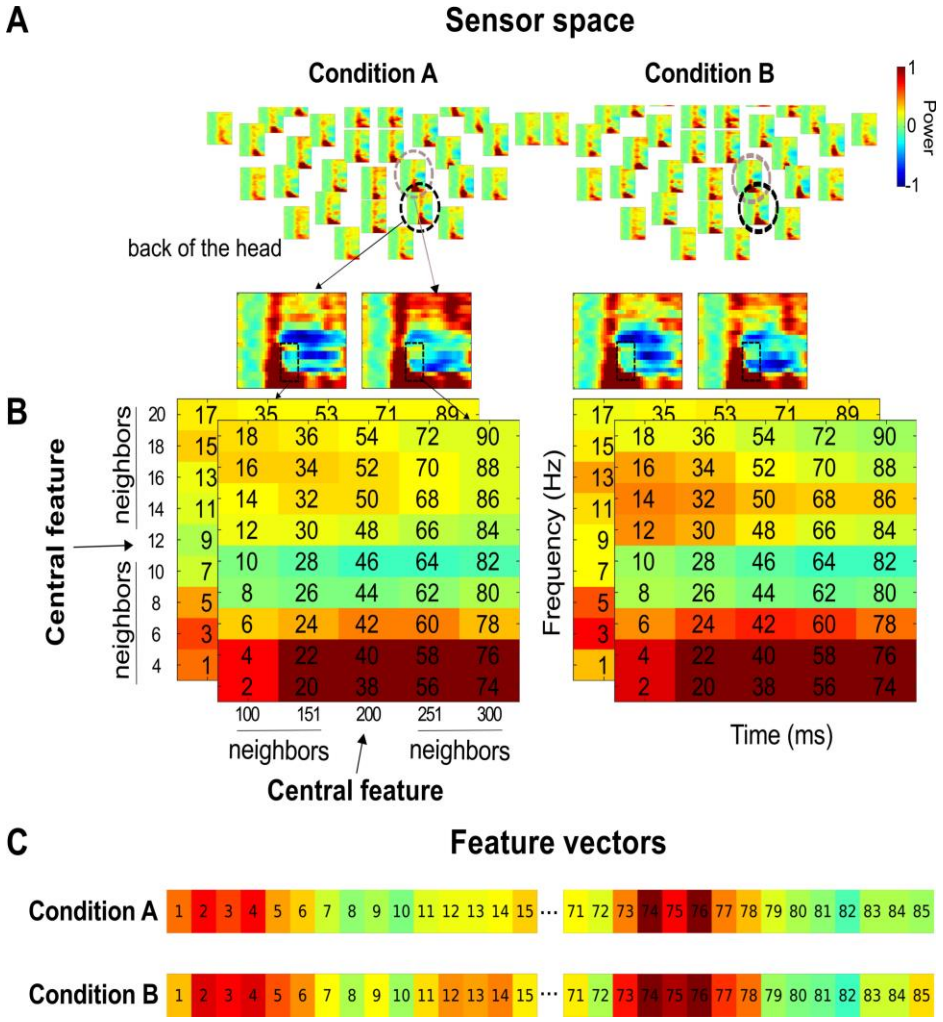


Figure 2.2. Feature selection. Schematic representation of the method we adopted for selecting the features used for the multivariate analysis. Here we show one specific step of the algorithm with the selected central sensor (black dotted circle) with one neighboring sensor only (gray dotted circle) for illustrative purpose. Panel A shows the time-frequency representations (colors indicate power intensity) in the posterior sensors of the MEG helmet in two conditions of interest (condition A and B). The arrows starting from the circles indicate the corresponding magnified sensors. Panel B shows enlarged views of the two example sensors for condition A and B. The dotted rectangles illustrate an example time-frequency bin (2 neighboring bins per side for the time dimension; 4 neighboring bins per side for the frequency dimension; see the Methods section for details). For feature selection, for each time-frequency bin, we scanned each individual sensor with its 10 neighboring sensors. Panel B shows a matrix representation of the specific sensor/frequency/time bins. We then rearranged the dimensions of the matrix from 3D to 1D to obtain the corresponding feature vectors for condition A and B (Panel C). The feature vectors were used as input for the decoding analysis over sensors, frequency, and time. Specifically, the feature vectors were partitioned in independent chunks and used for training and testing the classifier. In the depicted example, each feature within the matrices was assigned with a number to show the same feature within the feature vectors for visualization purposes.

MEG-MVPA demonstrates that LOTC encode abstract action representations earlier than frontal regions

To create subsets of trials to feed the classifier with the aim of differentiating between neural responses related to the observation of grasping and pointing actions irrespective of effector and reach direction, for each subject we divided the dataset into two independent halves, each containing only movements with a complementary combination of effector and reach direction. The first half included left hand movements to the right and right hand movements to the left, and the second half included left hand movements to the left and right hand movements to the right. We further divided the data into independent chunks, each of which contained at least 136 trials (depending on the number of trials remaining after artifact rejection) of a specific condition of interest. Then, for each half, we adopted a leave-one-chunk-out cross-validation method. We used 3 chunks associated to a specific condition for training, and a corresponding chunk with the complementary effector and direction for testing (cross-condition classification). This procedure was repeated for all chunks. Note that within a chunk the only dimension that differed across trials was the type of movement: grasping vs pointing. Thus, we assumed that the classifier learnt to discriminate between these two classes of stimuli. For example, if the training dataset contained the conditions grasping to the right with the right hand and pointing to the right with the right hand, the testing dataset contained the conditions grasping to the left with the left hand and pointing to the left with the left hand. For this type of classification, the classifier had to rely on differences between the two types of movements. If the model was able to discriminate between the two movements in the independent subset, this indicates that it had learnt the difference between the two types of movements using the previous training subset, generalizing across effector and reach direction. We adopted this approach for each possible factorial combination (for a complete list, see Table 1).

MEG-MVPA demonstrates that LOTC encode abstract action representations earlier than frontal regions

	Training: pointing vs grasping	*Testing: pointing vs grasping*
1	left hand, rightwards	right hand, leftwards
2	right hand, rightwards	left hand, leftwards
3	left hand, leftwards	right hand, rightwards
4	right hand, leftwards	left hand, rightwards

Table 2.1. Cross-comparisons used for training and testing. Classifiers were trained and tested in the following cross-comparisons: (1) training: distinguish between observed grasping and pointing actions performed with the left hand towards the left; testing: distinguish between observed grasping and pointing actions performed with the right hand towards the right; (2) training: distinguish between observed grasping and pointing actions performed with the right hand towards the right; testing: distinguish between observed grasping and pointing actions performed with the left hand towards the left; (3) training: distinguish between observed grasping and pointing actions performed with the right hand towards the left; testing: distinguish between observed grasping and pointing actions performed with the left hand towards the right; (4) training: distinguish between observed grasping and pointing actions performed with the left hand towards the right; testing: distinguish between observed grasping and pointing actions performed with the right hand towards the left. In this way, the classifiers could use information related to the type of movement only. The four accuracies determined with 1), 2), 3), and 4) were then averaged

The testing phase provided accuracy maps for each participant reflecting the classifiers' performance in discriminating between the two observed movements irrespective of effector and reach directions [in a similar way as traditional fMRI searchlights (Kriegeskorte et al., 2006), except that the features consist of (sensor, time, frequency)-triples rather than voxels]. We thus had information regarding where, when, and at which frequency band it was possible to distinguish between the abstract neural representations of the two movements.

In order to assess the reliability of the performance of the classifier, we used a non-parametric method (permutation test, similar to the procedure described above for the univariate analysis; see Maris and Oostenveld, 2007). In this case, we used the difference between the obtained classification accuracy and chance level accuracy (the accuracy expected under the null-hypothesis of no difference between the two conditions, meaning 50%) to compute the test statistic used in the permutation steps (see univariate method).

MEG-MVPA demonstrates that LOTC encode abstract action representations earlier than frontal regions

Any effect observed at sensor level has to be generated by neural sources. To visualize the sources underlying the cross-decoding effects for the frequency bands and time windows observed at the sensor level, we conducted a multivariate analysis at the source level, adopting the same searchlight approach as before (Kriegeskorte et al., 2006). Note that multivariate analysis was necessary here to identify which regions of the brain represented actions at an abstract level (i.e., generalizing across effector and reach direction). We reconstructed the source activity for the frequency bands and time windows that were significant at sensor level and expected to identify which regions of the brain were able to decode between grasping and pointing across effector and reach direction. We obtained estimates of frequency power at each grid point using a beamformer algorithm (see previous section) on a single trial basis. A searchlight was defined taking the power values at each grid point with its neighbours in a circle of 2 cm radius. For each participant, we found the accuracy maps indicating the performance of the classifier in discriminating between the two observed movements (irrespective of effector and reach direction). For descriptive purposes, we report the clusters showing the greatest classification accuracy.

Results

Behavioural experiment

We computed a chi-square test to evaluate at which time point participants' performance was significantly higher than chance level (50%). We found that performance of the participants was not different from chance level at 167 ms (X-squared = 11.7307, df = 13, p = 0.5498) and at 200 ms (X-squared = 21.4835, df = 13, p = 0.0639). Performance was significantly higher than chance level from 233 ms onwards (X-squared = 58.0318, df = 13, p = 1.178e-07). This means that participants were unable to distinguish the two actions if videos were shorter than 233 ms. Since mean movement onset in the videos (defined as the time point at which the rest position was released and hand preshaping was initiated; see Stimuli) was 191 ms (SD:

90 ms; median: 183 ms), this indicates that the two actions were perceptually indistinguishable before movement onset.

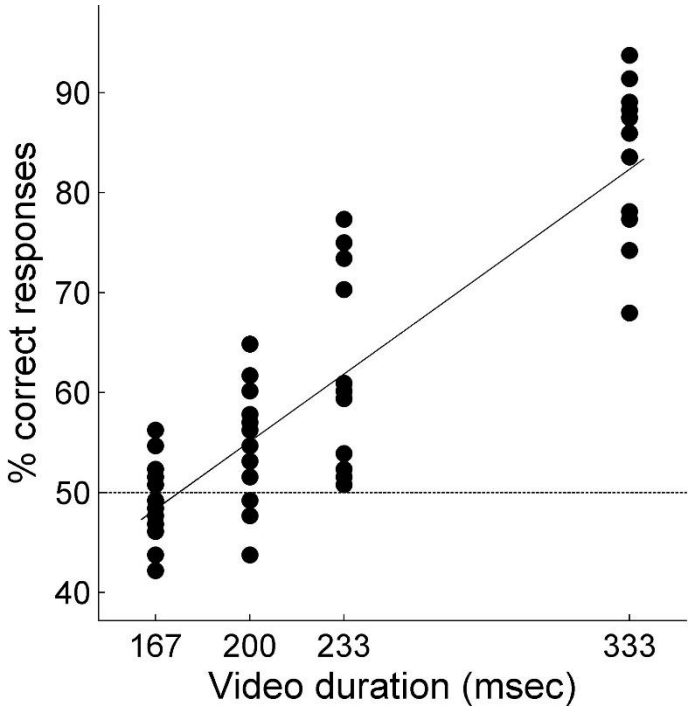


Figure 2.3. Behavioral results in action classification. Behavioral performance (% correct) for categorizing the two observed movements (grasping, pointing) as a function of video duration, collapsed across effector and reach direction. As expected, participants responded more accurately with increasing video duration. Statistical analysis confirmed that participants reached above chance performance in classifying the two movements from 233ms onwards (for details, see Material and Methods, Statistical Analysis: Behavioral Experiment). Each dot represents data from a single participant. The continuous line indicates the linear model that best fits the data.

MEG experiment

Univariate analysis. We first analysed the MEG signal using classical univariate methods to assess whether the stimuli induced a modulatory activity in the ongoing oscillations relative to rest. Low frequency bands such as alpha- and beta-bands are typically characterized by a decrease in power presumably due to neuronal activity synchronization in specific brain regions (Pfurtscheller & Lopes da Silva, 1999), indicating neural processing of the stimulus. Univariate analyses comparing the

activation period (after video onset) with the baseline (before video onset) demonstrated that passive observation of pointing and grasping modulates alpha- (8 – 12 Hz) and beta- (15 – 25 Hz) band power over posterior, parietal and frontal sensors, as well as the theta-band (4 - 7 Hz). **Figure 2.4A** shows one central sensor for illustrative purposes. In the depicted sensor, the alpha and beta rebounds related to post-observation processes are evident. Dotted lines approximately indicate the different stages of the movement (see Figure captions for details).

The decrease in power that we observed in the alpha and beta bands is in line with previous studies (Hari, 2006; Pineda, 2005) and has been suggested to reflect sensorimotor system activity. Further, the increase in power in the theta- and low alpha- (4– 8 Hz) band has been observed during memory tasks (Jensen & Tesche, 2002). In addition, modulation of these low frequencies has been reported during action observation, both in humans (Frenkel-Toledo et al., 2013; Pavlidou et al., 2014a, 2014b) and monkeys (Caggiano et al., 2014; Kilner et al., 2014).

MEG-MVPA demonstrates that LOTC encode abstract action representations earlier than frontal regions

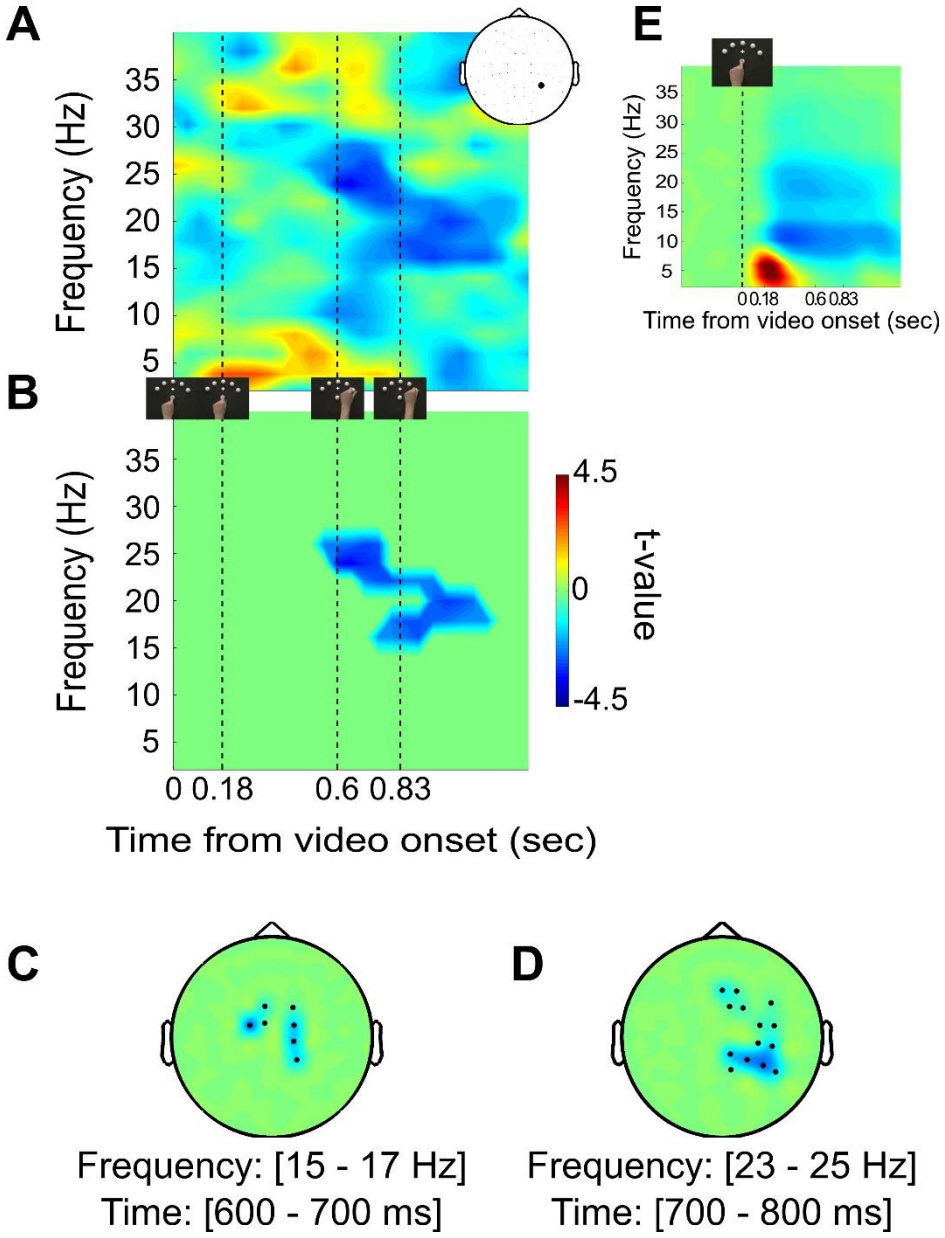


Figure 2.4. Theta-, alpha- and beta-band activity during action observation and univariate contrast.A, Time-frequency representation of the difference (expressed in t scores) between grasping and pointing (collapsed across effector and reach direction) for the sensor highlighted in the head model. The four dotted lines indicate the following events, from left to right: (1) video onset, (2) median movement onset, (3) approximate time at which the hand touches the object (around 550 ms), (4) video offset (833 ms). B, Same as A, but those time-frequency bins that did not survive the permutation test with Monte Carlo and cluster-based method for multiple comparisons correction were set to zero. **C, D**, Topography representation of the two frequency-bands observed in B. E, Power change during action observation relative to baseline (fixation cross) over a representative sensor. The power change was calculated as (activation – baseline)/baseline, such that 1 indicates 100% increase respect to baseline and -1 indicates 100% decrease respect to baseline. The classical power decrease in alpha and beta bands following observed movement onset (at t=0s) is evident.

MEG-MVPA demonstrates that LOTC encode abstract action representations earlier than frontal regions

A direct comparison of grasping and pointing movements (collapsing over effector and reach direction; see Materials and Methods, Univariate Analysis) showed a significant differential modulatory activity in beta (central frequency: 24 Hz) and alpha (central frequency: 16 Hz) band power over sensorimotor sensors at a late latency only (from around 750 to 1100 ms, and from around 500 to 750 ms, respectively). **Figure 2.4B** illustrates this effect for the same representative significant sensor as in **Figure 2.4A** over central regions. Bluish colours indicate that the power decrease is greater for grasping than for pointing; reddish colours indicate the opposite. **Figure 2.4C-D** shows the topography representations of the significant sensors in two selected subsets of frequency bands and time windows that were all located over central and right central sensors. These results show that (a) the brain processes underlying the two actions are different, and that (b) sensorimotor areas might be involved. The fact that grasping induces a greater decrease than pointing could be due to the higher complexity of this movement, which in turn is likely to recruit more neural resources. However, this differential activity appears quite late (at around 600 ms), long after the two movements are perceptually distinguishable. Thus, there must be another, earlier, process that allows for discriminating between the two movements, which the univariate analysis did not reveal.

Multivariate analysis. Figure 2.5A-C shows the results of the multivariate analysis at the sensor level. Two types of representations are provided: 1) a time-frequency representation, to show the dynamics of all the considered frequencies at each time point in a specific subset of sensors (panel A); 2) a topographical representation, to show the spatial information at specific time points and frequency bands (panel B, C). The inset of **Figure 2.5A** shows the two time-frequency clusters that survived the multiple comparisons correction. The lateral plots show the averaged t values over the sensors highlighted on the two topoplots in the middle. We observed that the classifier was able to significantly ($p<0.05$; corrected for multiple comparisons using a cluster-based method; maximum accuracy: 53.46%) discriminate between the two observed movements, generalizing across effector (left and right hand) and reach direction (left and right) over posterior sensors as early as 150 ms up until 550 ms in the low

alpha/theta range (see **Figure 2.5A**, left panel; for a direct comparison with univariate analysis, see **Figure 2.7**). By contrast, significant discrimination over more anterior sensors was possible only within a window of 550 – 1200 ms, i.e. at a late stage of the video, when the hand interacts with the object (see **Figure 2.5A**, right panel). **Figure 5B-C** show the topographies at different times and frequencies, selected according to the following criteria:

1) As times of interest, we selected the central point of the time windows selected based on the significant clusters that survived the significance test, i.e. 400 ms [200 - 600ms] for the cluster obtained in the earlier time window, and 900ms [600 - 1200ms] for the cluster obtained in the later time window.

2) Frequency bands were chosen based on previous studies showing a modulation of the low alpha (8-10 Hz) and high-theta (6-8 Hz) bands (Frenkel-Toledo et al., 2013) and the high alpha- (8-14 Hz) and beta- (15-25 Hz) bands during action observation (e.g. Pineda, 2005). For each time of interest (400ms and 900ms), we selected the peak frequency within each considered frequency band (i.e., 6, 8, 10, and 18Hz).

To examine the cortical sources of the effects shown in **Figure 2.5A-C**, we carried out another multivariate analysis at source level, adopting the same cross-comparisons as we did for the sensor analysis (see Materials and Methods for details). To find the sources at 400 ms for the frequencies 6Hz and 8Hz, we used temporal smoothing of 4Hz and time windows of 50ms to 650ms and 212ms to 587ms, respectively. **Figure 2.5D-E** shows the decoding accuracies of all the sources projected on surface template MNI brains, thresholded to retain only those voxels with the 10% of the highest accuracies (for the corresponding mean and individual decoding accuracies, see **Figure 2.6**). For the 6 Hz signal, the highest decoding accuracies were found bilaterally in the LOTC, extending into the inferior temporal gyrus and the superior temporal gyrus in the right hemisphere, and the left superior parietal cortex, extending into the inferior parietal cortex (**Figure 2.5D**, left panel; see **Table 2.2** for MNI coordinates of the peak voxel in each cluster). The highest decoding accuracies for the 8Hz signal were located

in the left LOTC (**Figure 2.5D**, right panel), slightly anterior to the source identified at 6Hz.

Regarding the sources related to the decoding obtained in the late time window, we chose 900 ms as time of interest for the frequencies 10Hz and 18Hz (time windows: 600-1200ms and 678-1122ms, respectively; smoothing: 3Hz). For the 10 Hz signal, we obtained the highest decoding accuracies in right precentral gyrus (**Figure 2.5E**, left panel). For the 18Hz signal, we obtained highest decoding accuracies in the right inferior frontal gyrus (**Figure 2.5E**, right panel).

MEG-MVPA demonstrates that LOTC encode abstract action representations earlier than frontal regions

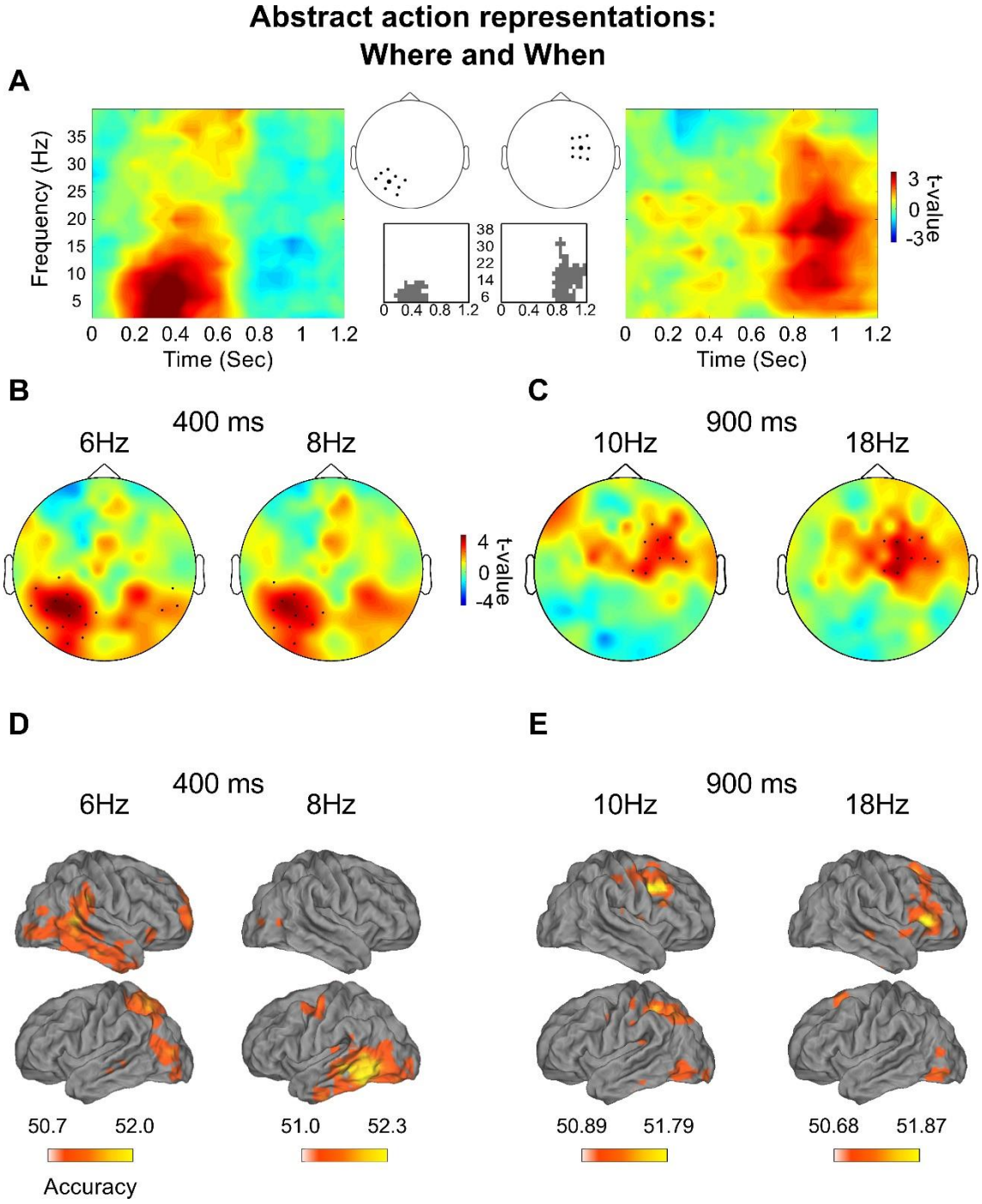


Figure 2.5 Results of the neural spatiotemporal decoding. To identify abstract action representations of the observed actions (e.g. observing "grasping" irrespective of whether it was performed with the left or the right hand), we trained the MVPA classifier to discriminate between pointing and grasping using one effector (e.g. the left hand) and one reach direction (e.g. towards the left), and tested the performance of the classifier using an independent data set, using pointing and grasping movements performed with the other hand towards the opposite reach direction. We decoded the observed movements over time bins, frequency bins and sensors using a time-frequency-channel searchlight analysis. **A**, The lateral plots show the time-frequency representation of the decoding in sensors depicted in the inset topoplots. Reddish colours indicate higher classification.

Sensors were selected on the basis of the highest decoding accuracy at the frequency of interest. The central inset shows the two clusters that survived the correction for multiple comparisons (cluster obtained at early time point: 200 to 600ms; cluster obtained at late time point: 600 to 1200ms). **B**, Topography of the decoding at 400 ms and low frequencies (6Hz and 8Hz; smoothing: 4Hz). **C**, Topography of the decoding at 900 ms and higher frequencies (10Hz and 18Hz; smoothing: 3Hz). **D-E**, Sources accounting for the decoding effect found at sensor level, thresholded to retain only those voxels with the 10% highest decoding accuracies. For sensor level analysis only, significant differences were computed using permutation analysis and Monte Carlo methods and results are cluster corrected for multiple comparisons. Maps were projected on the PALS altas (Van Essen, 2005), using Caret software (Van Essen et al., 2001).

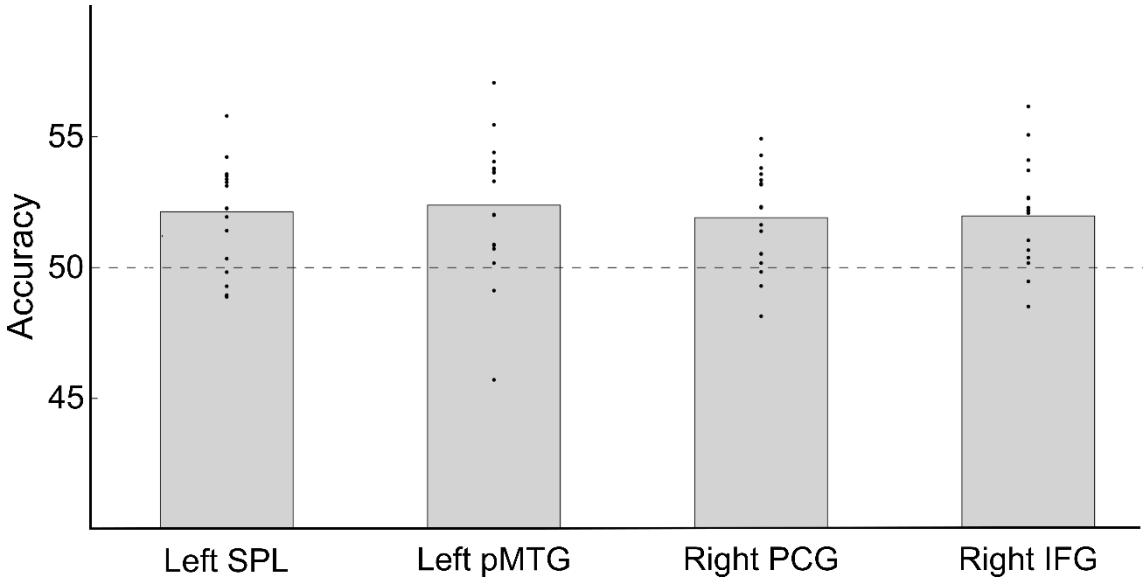


Figure 2.6. Maximum accuracy within each region. Within each identified source, the voxel with the maximum mean accuracy was selected and plotted with individual accuracies (black dots). Left MTG: Middle Temporal Gyrus (MNI: -50 -64 12); Left SPL: Superior Parietal Lobule (MNI: -20 -56 48); Right PCG: Precentral Gyrus (MNI: 28 -6 28); Right IFG: Inferior Frontal Gyrus (MNI: 20 24 28). Refer also to Table 2.

MEG-MVPA demonstrates that LOTC encode abstract action representations earlier than frontal regions

Area	*Frequency (Hz)*	*Time (msec)*	*X*	*Y*	*Z*
Left pITG	6	200	-52	-56	-12
Left SPL	6	400	-20	-56	48
Left pMTG	8	400	-50	-64	12
Right PCG	10	900	28	-6	28
Right IFG	18	900	20	24	28

Table 2.2. MNI coordinates of sources. MNI coordinates of the sources (clusters) found in the two different frequency bands, with the respective labels taken from the Anatomical Automatic Labeling (AAL) database (Tzourio-Mazoyer et al., 2002). **pITG**: posterior portion of the Inferior Temporal Gyrus; **pMTG:** posterior portion of the Middle Temporal Gyrus; **SPL:** Superior Parietal Lobule; **IFG**: Inferior Frontal Gyrus (triangular part); **PCG:** Precentral Gyrus. **Table 2. MNI coordinates of the sources.** MNI coordinates of the sources (clusters) found in the two different frequency bands, with the respective labels taken from the Anatomical Automatic Labeling (AAL) database (Tzourio-Mazoyer et al., 2002). **pITG**: posterior portion of the Inferior Temporal Gyrus; **pMTG:** posterior portion of the Middle Temporal Gyrus; **SPL:** Superior Parietal Lobule; **IFG**: Inferior Frontal Gyrus (triangular part); **PCG:** Precentral Gyrus.

MEG-MVPA demonstrates that LOTC encode abstract action representations earlier than frontal regions

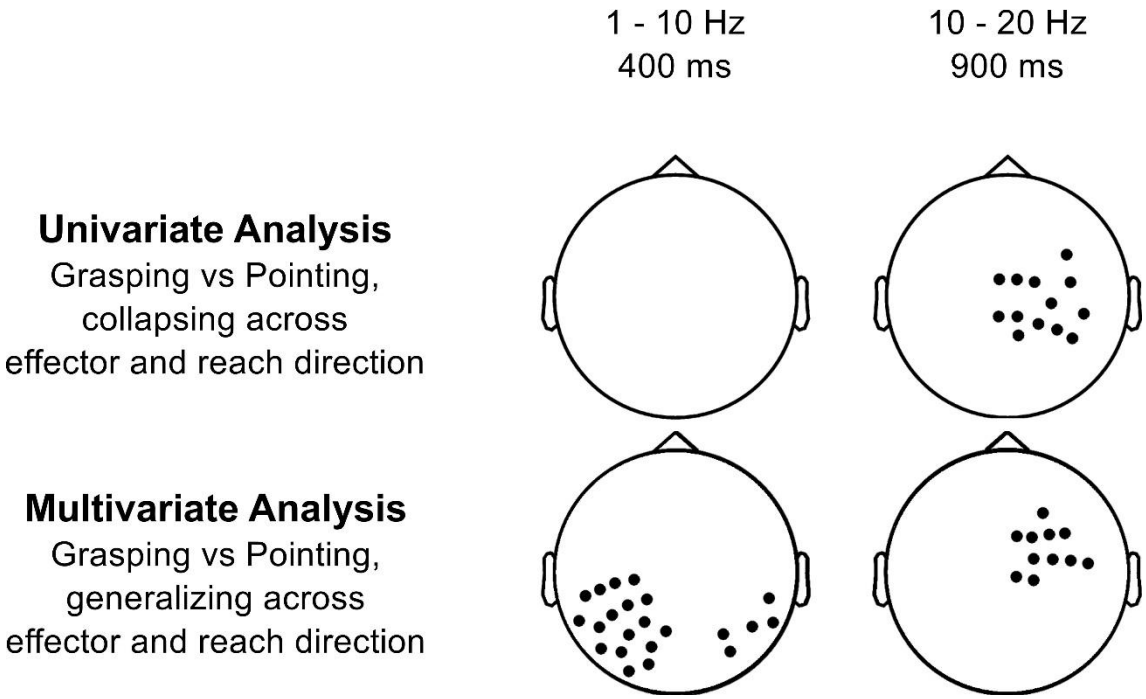


Figure 2.7. Comparison between univariate and multivariate analyses. Comparison between univariate (top row) and multivariate (bottom row) analyses in two time windows ([200 – 600ms] and [600 – 1200ms]). The upper topoplots show the sensors that survived the permutation test when comparing grasping vs pointing (collapsing across effector and reach direction). The lower topoplots show the sensors that survived the permutation test when comparing the observed accuracy of the classifier to distinguish between pointing and grasping (generalizing across effector and reach direction) against chance level (50%). Multivariate analysis was more sensitive in detecting the subtle differences between the neural signals induced by observation of the two movement types in the earlier time window. All shown clusters are corrected for multiple comparisons ($p<0.05$).

MEG-MVPA demonstrates that LOTC encode abstract action representations earlier than frontal regions

To show a complete overview of the temporal dynamics of the neural decoding at sensor space, we plotted the decoding accuracy (expressed in t values) for separate time bins (50-150, 150-250, 250-350, 350-450, 450-550, 550-650 ms for the early observed decoding, **Figure 2.8A**; 350-450, 450-550, 550-650, 650-750, 750-850, 850-950, 950-110 ms for the late observed decoding, **Figure 2.8B**), averaged across frequency bands (theta: 2-6 Hz; low alpha: 7-9; alpha: 9-11; beta: 17-19). This figure shows how the effect over posterior sensors evolves over time, and that anterior sensors do not show up before around 700 ms.

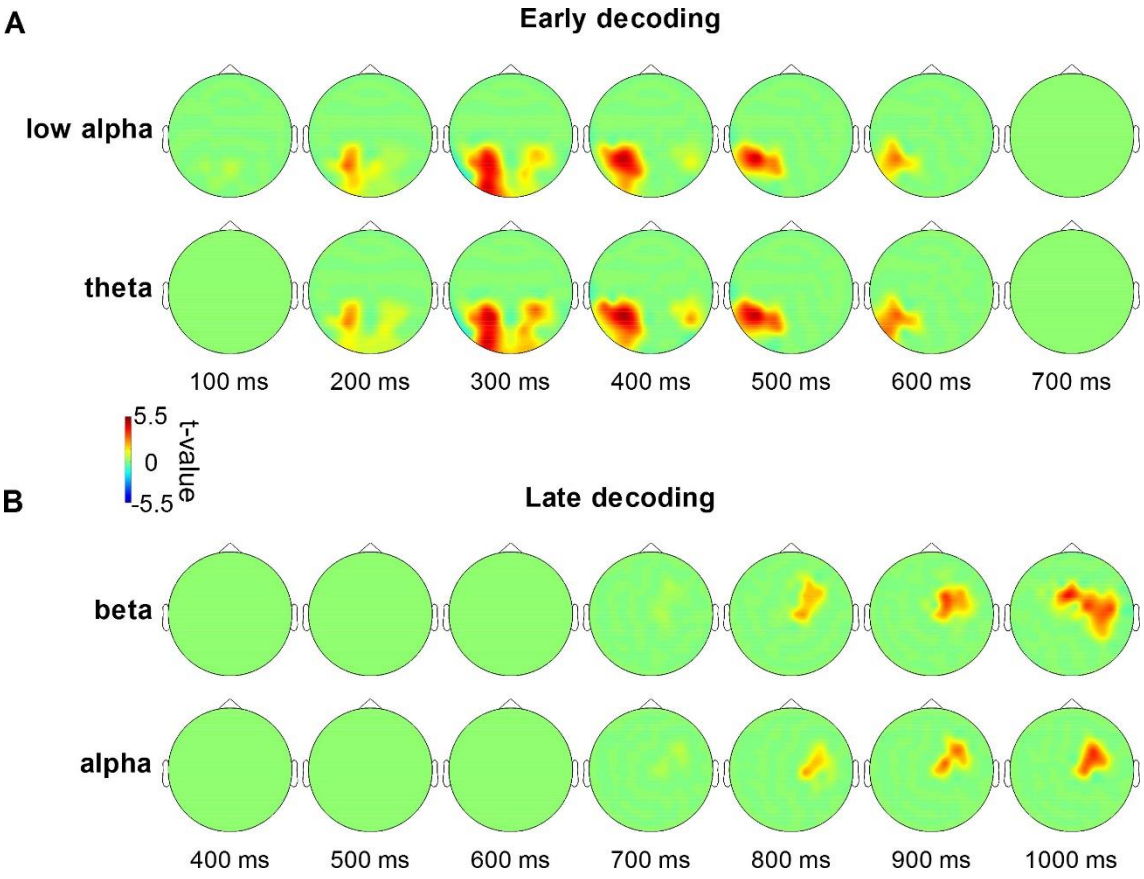


Figure 2.8. Neural decoding over time. The topoplots show the dynamics of above-chance accuracy (expressed as t scores) of the classifier in discriminating observed grasping and pointing (generalizing across effector and reach direction) for specific frequency bands (theta: 5-7 Hz; low alpha: 7-9 Hz; alpha: 9-11 Hz; beta: 17-19 Hz). The earliest significant decoding occurs in the posterior part of the configuration helmet in the lower frequency bands.

MEG-MVPA demonstrates that LOTC encode abstract action representations earlier than frontal regions

To further evaluate the reliability of the classifier, we also used a simulation approach. Specifically, we ran a Monte Carlo simulation to estimate the probability of finding an accuracy of 53.46% under the null hypothesis of chance accuracy. The cross-validation partitioning scheme divided the data into two independent halves (see Table 1 and Methods), with the first half containing left hand rightwards and right hand leftwards trials, and the second half containing right hand rightwards and left hand leftwards trials. In each independent half, there were two folds, with a minimum of 136 trials (across participants and halves) after rejecting trials with artifacts and balancing the partitions so that each of the two actions occurred equally often. For each participant separately, we found that the correlation of classification accuracies for the test sets in two folds to be r=0.3289 (median across participants and the two independent halves). Thus, in our simulation we used the same value as follows. For each permutation, uniformly distributed (on the interval [0, 1]) random data was generated for two independent halves, two folds, 136 samples, 17 participants. To assess the effect of dependency we used 3 sets of independently normally distributed data i_1, i_2 and i_{common}. To match the correlation between accuracies, for each independent half of the data, data was made dependent through $d_1 = i_1 * \gamma + i_{common} * (1- \gamma)$ and $d_2 = i_2 * \gamma + i_{common} * (1- \gamma)$, with $\gamma=0.415$ found through binary search to match the correlation (r=0.3289) across dependent folds as observed in the original data. For each iteration, classification accuracy was simulated by dividing the number of samples that exceeded 0.5 in d_1 and d_2 by the number of samples. 0.5 was subtracted to obtain classification accuracies relative to chance.

To assess the effect of independence, we also ran the same analysis setting $\gamma=0$ (corresponding to r=0, i.e. full independence between folds), and $\gamma=1$ (corresponding to r=1, i.e. full dependence between folds).

We used 100,000 iterations and found that the maximum classification accuracies found in the data (using r=0.3289 for fold correlation) was significant, $P_{MC,sensor, r=0.3289}<0.00001$; for the latter, no iteration showed a higher mean than that observed in the data (**Figure 2.9**). We obtained similar results for the additionally simulated

cases of fully independent folds (r=0), $p_{sensor,\ r=0.00}<0.00001$, and dependent folds (r=1), $p_{sensor,\ r=1.00}<0.00001$.

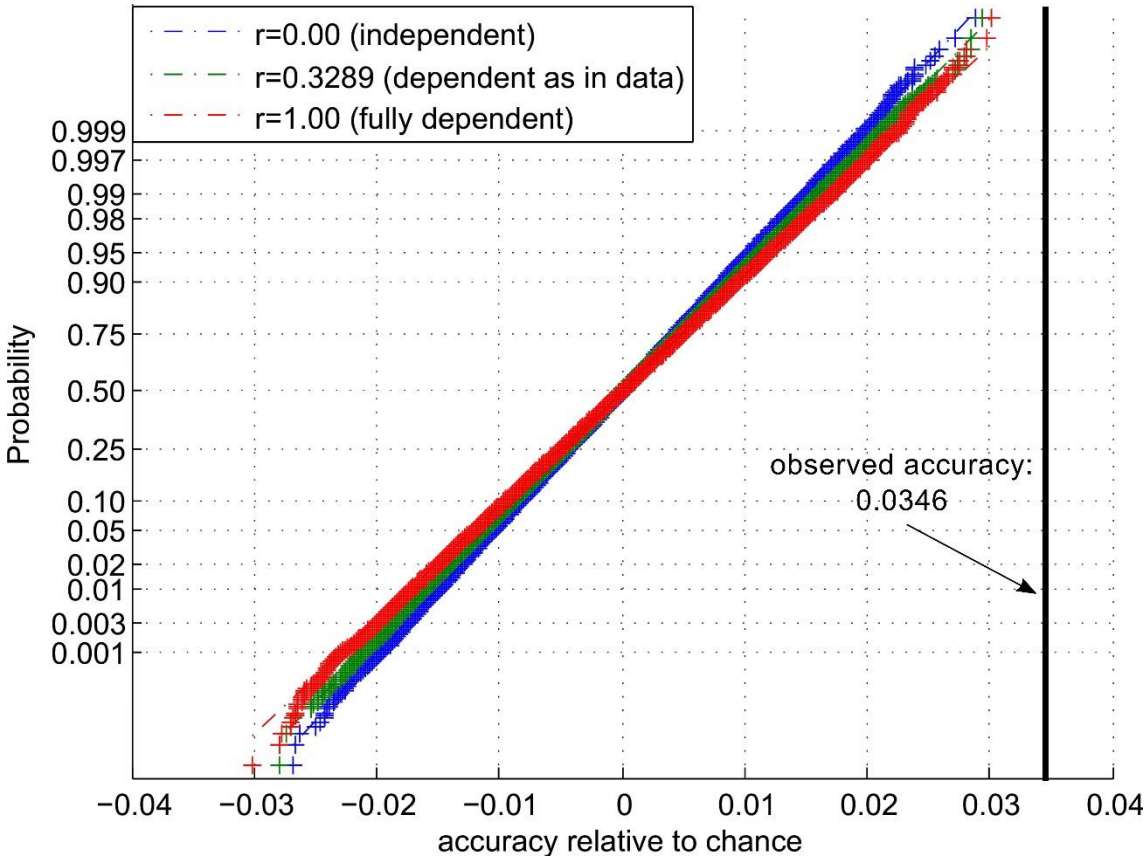


Figure 2.9. Simulation analysis. Illustration how 'low' classification accuracy (53.46% for sensor data; 50% is chance level) can be highly significant, using normal distribution probability plots of Monte Carlo simulated classification accuracy distribution (relative to chance, 50%). The simulation uses the same parameters as used in the study (17 participants, minimum after trial rejection 544 trials per participant, same cross-validation scheme as used in original data). Dependency across cross-validation folds was set to r=0.3289 (green crosses) to match the value observed in the original data; for comparison, also results are shown for the cases of no dependence (r=0.00; blue) and full dependence (r=1.00; orange). The maximum classification accuracy above chance as observed in the original data is indicated by a black line.

MEG-MVPA demonstrates that LOTC encode abstract action representations earlier than frontal regions

Discussions

Using MVPA of MEG data, we found that LOTC has the earliest access to abstract action representations. By contrast, precentral regions, though recruited relatively early, have access to abstract action representations substantially later than LOTC. Behavioral data indicated that participants were not able to distinguish between the two actions before 233 ms, and this latency is comparable with the one observed in LOTC.

Early abstract action representations in occipito-temporal and parietal regions

Although MEG has a lower spatial resolution than fMRI, we can confidently say, based on the topographical results and source analysis, that the source that accounted for the decoding effect we found at the early stage was located within the left and right LOTC. LOTC hosts regions sensitive to body parts, kinematics, body postures, manipulable objects, and observed movements (Buxbaum et al., 2014; Downing & Peelen, 2011; Lingnau & Downing, 2015; Pavlidou et al., 2014a, 2014b; Valyear & Culham, 2010). LOTC has been shown to be modulated when participants are required to process the meaning, in comparison to the effector, involved in an action (Lingnau & Petris, 2013). Moreover, LOTC is recruited during the semantic processing of verbs (e.g. Papeo et al., 2014), and lesions to this region are associated with impairments in action recognition (Kalénine et al., 2010; Urgesi et al., 2014). In line with this view, a recent lesion study demonstrated that lesions to primary motor, somatosensory and inferior parietal lobule were accompanied by impaired action performance. By contrast, lesions to posterior LOTC were associated with impaired action recognition, whereas lesions to anterior LOTC were accompanied by impairments in both tasks (Tarhan et al., 2015). Taken together, these studies suggest that LOTC is well suited to integrate various sources of information that are crucial for action understanding.

Neuroimaging studies using MPVA of fMRI data have recently shown that LOTC also contains abstract representations of observed actions, e.g. action representations that generalize from action execution to action observation and vice versa (Oosterhof et al., 2010), that generalize across viewpoint (first person, third person; see Oosterhof et al.,

2012a), kinematics (Wurm & Lingnau, 2015), and the object involved in the action (Wurm & Lingnau, 2015; Wurm, Ariani, Greenlee, & Lingnau, 2015). Importantly, our study shows that such abstract representations are available before observing this kind of representation in precentral regions, around the time when there is enough information in the stimuli to distinguish between the two types of actions. Our findings are compatible with cognitive theories of action understanding that predict the earliest encoding of the meaning of an action outside the motor system. By contrast, our results are not compatible with motor theories of action understanding that would predict the earliest access to abstract action representations in precentral regions.

The fact that we observed abstract action representations in LOTC earlier than in precentral regions is compatible with a framework suggested by Kilner (2011). According to this view, the middle temporal gyrus (MTG) in the LOTC and the anterior portion of the IFG (aIFG) encode the most likely goal or intention of an action (e.g. grasping an object), which is communicated to the posterior portion of the IFG, where the most likely action is selected. In this framework, the role of the posterior IFG would be to generate a concrete instance of the action (e.g. grasping an object on the left using the right hand) through motor simulation. In contrast to motor theories of action understanding, the role of this motor simulation would not be to provide access to the meaning of the action, but rather to contribute to the generation of the predicted sensory consequences of the most likely action.

We observed abstract action representations at around 400 ms in the left SPL as well, extending into the inferior parietal lobule (IPL). This result is in line with previous monkey (Fogassi et al., 2005; Rizzolatti et al., 2014) and human fMRI studies (Grafton & Hamilton, 2007; Leshinskaya & Caramazza, 2015; Oosterhof et al., 2010, 2012b; Wurm & Lingnau, 2015; Wurm et al., 2015) suggesting that, similar to LOTC, this region contains abstract action representations. The observation that IPL/SPL has access to abstract action earlier than precentral regions, raises the possibility that this region might play an intermediate role between LOTC and precentral regions (see also Wurm et al., 2015). In line with this view, Pavlidou et al (2014b) demonstrated that the

difference between plausible and implausible actions is first obtained over left temporal sensors, followed by parieto-occipital and sensorimotor sensors.

Late abstract action representations in precentral regions

The contrast *observation vs baseline* showed a modulation of the high alpha and beta frequency bands over central sensors during passive action observation (**Figure 2.4E**), an effect that has been suggested to be related to sensorimotor processing in motor and premotor regions (Pineda, 2005). Although we observed an early modulation of high alpha and beta frequencies in precentral regions for observation versus baseline, these regions had access to abstract representations of the observed actions substantially later than the time at which the actions were distinguishable. This finding makes a determinant role of precentral regions in action understanding implausible. In line with this view, damage to precentral regions does not necessarily impair the ability to understand actions (Negri et al., 2007; Kalénine et al., 2010; but see Pazzaglia et al., 2008). If precentral regions do not play a determinant role in action understanding, what could be the alternative role of the late abstract action representations we obtained in these regions? Since LOTC and precentral regions are functionally interconnected (Engel et al., 2013; Kilner, 2011; Nelissen et al., 2011; Papeo et al., 2014; Turken & Dronkers, 2011), higher-level representations in precentral regions have been suggested to be a result of information spreading throughout the network (Mahon & Caramazza, 2008). Instead of providing access to the meaning of an action, precentral regions thus might be recruited to plan an appropriate movement in response to the observed action as a consequence or in parallel to the process of action understanding.

Potential caveats

One potential limitation regarding the interpretation of our results is related with the fact that one of the main distinctions between pointing and grasping, next to the pre-shaping of the hand, is the number of fingers involved. It is therefore difficult to disentangle whether our classification is based on the number of fingers involved in the movement, the pre-shaping of the hand while approaching the target, or a

combination of the two. Note that pointing and grasping movements are defined both by the number of fingers involved and by the hand configuration; in other words, understanding actions could rely on the number of used fingers as well as on an examination of the pre-shaping of the hand.

Another possible criticism could be that we were able to distinguish between the two movements based on the MEG signal as early as 150 ms, which seems counterintuitive given that the mean movement onset in the videos was around 191 ms. There are several not mutually exclusive explanations for this observation. First, movements started before 150 ms in 43.8% of the videos (see Material and Methods). By contrast, the peak of decoding from the MEG signal was obtained at around 300 ms. Second, we had to apply a certain amount of temporal smoothing during time-frequency computation and during the searchlight analysis (see Materials and Methods). Consequently, when the algorithm analyzes the time bin at 150 ms, it also considers information present at 200 and 250 ms, which contained more information about movement type. This means that the absolute latency at which the two actions can be distinguished based on the MEG signal has to be interpreted with a grain of salt. Importantly, we do not aim to draw strong conclusions regarding the *absolute* onset at which movements can be decoded in the different regions, but rather about the *relative* difference between putative regions involved in action understanding. Thus, our conclusion still holds: LOTC encodes abstract representation of actions earlier than precentral regions.

One might argue that although we observed the strongest source in the early time window within LOTC, the source analysis also revealed a small left frontal region. This frontal source is very likely generated by a single temporal source, in line with the observation that no frontal sensors showed significant decoding in this early time window (**Figure 2.8**). Note that the absence of a frontal source in the early time window does not proof that such a source does not exist. What we can state with a certain confidence, though, is that the same analysis that revealed a strong and reliable source in LOTC did not reveal any frontal source in the early time window.

MEG-MVPA demonstrates that LOTC encode abstract action representations earlier than frontal regions

Conclusion

Our results demonstrate that LOTC has access to abstract action representations substantially earlier than precentral regions, in line with the idea that action understanding occurs outside the motor system, with subsequent activation of precentral regions due to information provided from LOTC. Our results therefore provide important constraints for biologically plausible models of action understanding.

Chapter 3: Early abstract and concrete action representations in LOTC

Adapted from:

Tucciarelli R.,Weisz N., and Lingnau A.(in preparation). *Neural decoding of oscillatory patterns during action observation: investigating effector-independent and effector-dependent action representations.*

Introduction

The study described in the previous chapter (chapter 2) revealed that lateral occipito-temporal regions (LOTC) have access to abstract action representations in a time window immediately following the initiation of the observed action. Specifically, action-related information that is independent of the specific action implementation is encoded in LOTC. Crucially, this abstract information was available earlier in temporal regions than in frontal regions. We interpreted these results as evidence that LOTC plays a fundamental role in action understanding.

Our results and interpretation are in not line with the motor theory of action recognition (Pulvermüller et al., 2005; Rizzolatti et al., 2001), which would have predicted motor-related regions to have the earliest access to abstract action representations in view of the fact that involvement of brain areas in action comprehension "[…] occur early after stimulus presentation (e.g. within 200-300 ms)." (Hauk et al., 2008). Our result can instead be explained by cognitive theories which claim that action understanding occurs outside the motor system (Caramazza et al.,

2014; Mahon & Caramazza, 2008), more specifically in higher visual regions (Hickok, 2012; Lingnau & Downing, 2015).

Study 1 raises two fundamental questions: 1) how *selective* are the temporal and frontal regions that showed abstract action encoding? Specifically, are other types of actions, aside from grasping and pointing, also represented in temporal and frontal regions in an abstract fashion and with a similar temporal profile? In addition, 2) Do temporal and frontal areas also show a generalization from hand- to foot-related actions (and vice versa), and to which degree do the underlying neural dynamics differ from effector-specific representations?

To follow up on these questions, we designed a new MEG study. In order to test if frontal regions encode actions in general and not only pointing vs grasping actions, we used four new actions: *throwing*, *squeezing*, *kicking*, and *crushing*. Notably, the first two actions are performed with the hand, whereas the latter two are performed with the foot. Nevertheless, the four actions are at some level comparable across effectors: *squeezing with the hand* and *crushing with the foot* are conceptually similar, as in both cases the goal of the action is *to compress and thus destroy an object*; also *throwing with the hand* and *kicking with the foot* are conceptually similar, as in both cases one is *moving the object away from his own body*. Thus, the different types of effectors involved allowed us to investigate the level of abstractness encoded in a certain region. An area encoding concrete information would be able to discriminate only within-effector actions; on the other hand, a region containing abstract action information would be able to discriminate between conceptually similar actions across effectors. Importantly, we used a high variety of stimuli (i.e. action implementations) to control for low-level features. In this way, we were able to rule out that actions were decoded on the basis of low-level properties of specific action exemplars, like the number of fingers used, the object involved, etc.

To anticipate our results, in contrast to the predictions of motor theories of action understanding, we observed that during an early time window, following stimulus onset, most of the action information is encoded in lateral occipital regions, regardless

of whether the targeted information is abstract or concrete. A different pattern of results was observed at a later time window, with more concrete actions being represented also in frontal and prefrontal regions, notably also in the low-level areas specific for the effector used. Abstract information was also encoded in parietal regions at this late latency.

Materials and Methods

Participants

Twenty students from the University of Trento with normal or corrected-to-normal visual acuity and with no neurological disorders took part in the MEG experiment. All participants received a reimbursement of €25 at the end of the MEG session. All of them gave informed consent in accordance with the Declaration of Helsinki. The experimental procedures were approved by the Ethics Committee for research involving human participants at the University of Trento.

Stimuli

Stimuli (**Figure 3.1**) consisted of still images of either actions or human bodies (only the torso). The actions consisted of a wide collection of either daily (squeezing a lemon, crushing a box) or sports-related actions (throwing a rugby ball, kicking a football) performed with either the hands or the feet. Despite the great variety of specific actions included, they could be classified into four subordinate categories (**Figure 3.2**): squeezing and throwing (hand actions); crushing and kicking (foot actions). At a higher level, these actions could also be divided into two main categories: moving something away from the body (e.g. throwing a baseball ball) and compressing something (e.g. squeezing a sponge).

The images were either taken from the internet database Shutterstock (www.shutterstock.com) or were generated in-house using a Canon 5D Mark camera. Colours were reduced to grayscale, and the size cropped to 400x300 pixels using Gimp 2 (www.gimp.org). We took great care in choosing the images such that they varied as

much as possible with respect to low-level features (arm and leg positions, object positions, etc.) and contexts. This is because we were interested in abstract action representations irrespective of the ways in which these actions were implemented. We initially selected 24 exemplars each for the four categories. We then asked 15 (mean age ± std: 23.13 ± 2.53) people to rate how dynamic each action appeared on a scale of 1 to 6 and selected the 15 most highly rated action exemplars in each category.

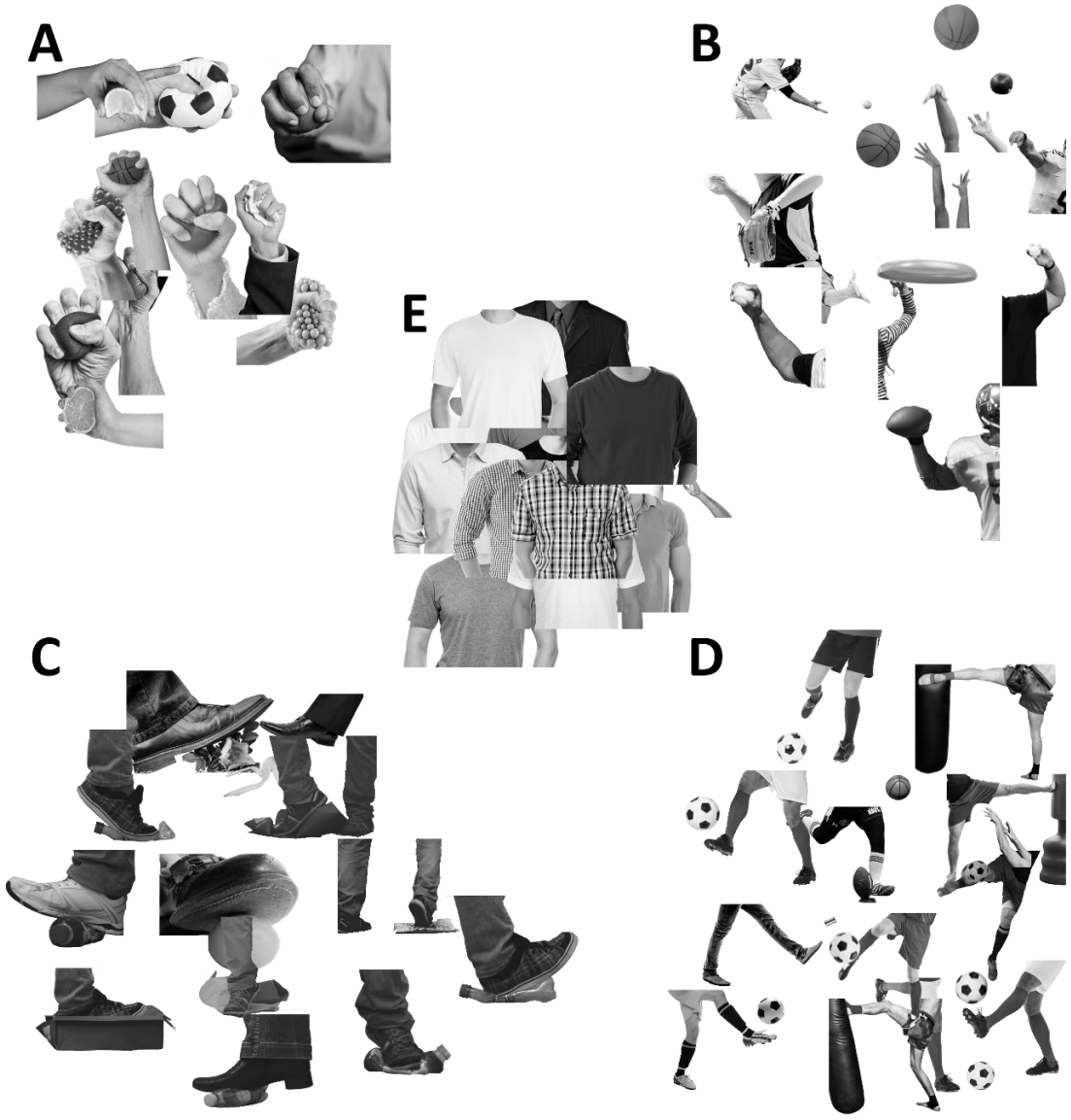


Figure 3.1. Stimuli. Stimuli (images) used in the experiment organized according to the effector (hand, **A,B** and foot, **C, D**) and type of action (squeezing, **A**; throwing, **B**; crushing, **C;** kicking, **D**). We used torso images as control stimuli (panel **E**). The high variety of action implementations (12 exemplars per action/body parts) guaranteed that the classifier could not learn to discriminate between actions according to low-level features.

The images were presented to the participants inside the MEG room through a projector (Vpixx Propixx) onto a translucent rectangular screen (1280x1024 pixels screen resolution; grey background; 60 Hz refresh rate). Stimuli were sent to the projector through the device Vpixx Datapixx to ensure time synchronization. We controlled stimulus presentation with ASF (Schwarzbach, 2011), a Matlab toolbox based on Psychtoolbox (Brainard, 1997).

Experimental procedure and design

Each single trial (**Figure 3.3**) consisted of the following events: **1)** a green cross for 1 second (blink phase); **2)** a white cross for 2-2.5 seconds (baseline period); **3)** an experimental stimulus for 0.8 sec (image phase); **4)** a white cross for 1 second (resting period). An MEG session consisted of eight runs. During each run, participants performed 75 experimental trials and 8 catch trials repeating the stimulus of the preceding trial. The total number of experimental trials per participant was 600 (75 trials by 8 runs). Each category (the four actions and the bodies) was repeated 120 times (15 exemplars by 8 runs).

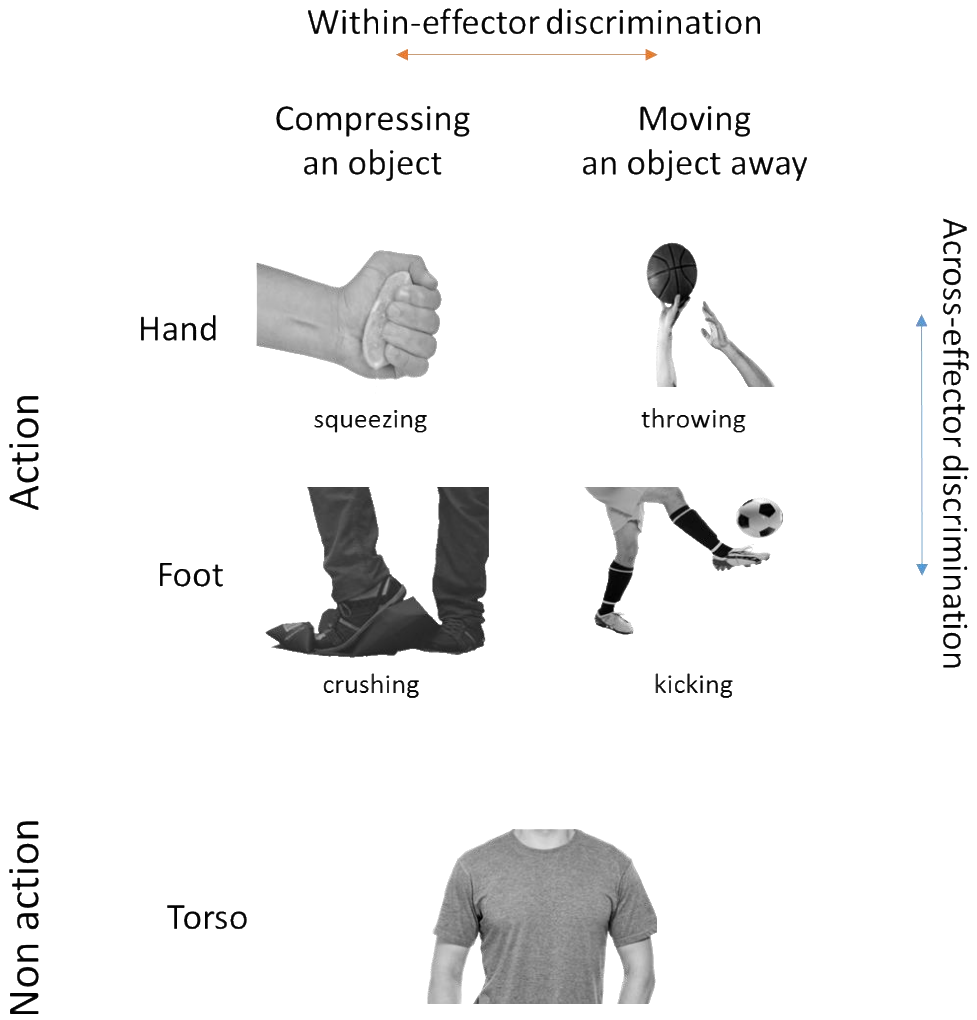


Figure 3.2. Stimuli and MVP analysis. To individuate neural patterns elicited during observation of within-effector and across-effector actions, we conducted a sensor-based searchlight multivariate (MVP) analysis based on a support vector machine (SVM) algorithm using CoSMoMVPA (Oosterhof et al., in preparation). Time-frequency patterns containing information related to the type of actions should be similar when the same action is observed and dissimilar when a different action is observed. We also tested information related to the type of effector used (within-effector discrimination) or independent of the effector (across-effector discrimination). We adopted a *leave-one-chunk-out* cross-validation approach. Specifically, for the within-effector analysis, a classifier was trained to discriminate between effector-specific actions (e.g. hand actions) and tested on actions performed with the same effector using trials from an independent dataset. For the across-effector analysis, we trained the classifier to discriminate actions performed either with the hand or the foot, and tested it using actions taken from a separate dataset.

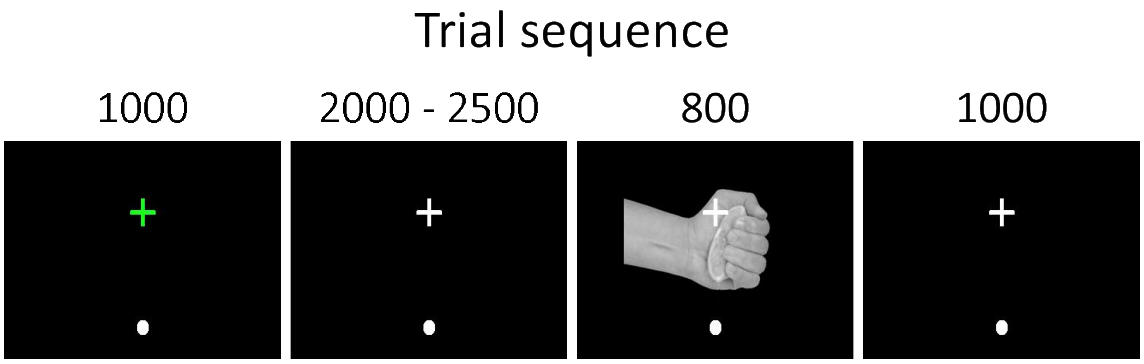


Figure 3.3. Example trial sequence. A trial started with a green cross (1000 ms), indicating to the participant to blink, followed by a white cross (2000-2500 ms), indicating that a stimulus was about to appear. Then, the stimulus (an image) was presented for 800 ms, along with the white cross (to maintain fixation). The trial ended with a white cross presented for 1000 ms. Participants had to attend to the images and report catch trials (repetition of the same image twice in a row) by making a saccade to the white dot at the bottom of the screen and immediately returning to the fixation position.

The total duration of a trial varied from 4.8 to 5.3 seconds (mean 5.05 sec) depending on the duration of the baseline period. Each run lasted around 7 minutes and each session about an hour. Between runs, participants were allowed to briefly rest before starting a new run. A longer break was typically taken after the fourth run.

Task

Participants were instructed to attentively observe the images presented on the screen. They were asked not to move and to fixate the white cross for the entire duration of the experiment. To ensure that they attended the images for the entire run, they had to report each time an identical image was repeated (catch trials) which occurred in 10% of the trials. In each run, any specific image could be repeated only once. We measured eye movements using an MEG-compatible eye-tracker (OEM eye tracker, SMI; 60 Hz sampling rate), and responses were given by a vertical saccade to a small dot below the image. After the saccade was performed and the response was reported, the participant had to return his gaze to the fixation cross. We decided to use eye movements to record the behavioural performance because we did not want any of the effectors of interest (foot and hand) being involved in the response. This

avoided any potential effect due to motor system activation (e.g. using only the hand to respond could have made the hand motor representation more active than the foot motor representation; using both effectors to respond could have made the task too confusing).

Data acquisition

Before entering the MEG room, one of the experimenters digitalized the head shape of the participant using a Polhemus system (Polhemus, Colchester, VT). We also placed three electrical coils at the participant's forehead and two behind the ears to acquire the head position inside the MEG helmet. All metal and magnetic materials were removed from the participant before entering the shielded room.

We measured neuromagnetic brain activity using a 306-channel (204 planar gradiometers and 102 magnetometers) whole head MEG system (Neuromag Elekta Oy, Helsinki Finland) at a sampling rate of 1000 Hz. Triggers were sent at stimulus onset and were used to create the epochs for data analysis.

Data analysis

MEG preprocessing

Data were preprocessed using Fieldtrip (Oostenveld et al., 2011). Preprocessing consisted of: **1)** high-pass filtering (1 Hz) using a two pass (forward and reverse) Butterworth IIR filter; **2)** parsing the continuous data into epochs (from 1.6 before to 1.6 after stimulus onset based on the triggers); **3)** downsampling each epoch from 1000 Hz to 350 Hz; **4)** appending the eight runs to have a single dataset per participant.

Artefact rejection

Artefact rejection consisted of three steps: 1) clearly noisy trials and channels were initially selected based on global visual inspection of each channel and trial; for each channel, we also computed its variance with respect to the mean signal across trials and its max and min values; these statistics, when considered independently of other channels, are good indicators of noisy activity. We did the same for each trial (we

computed statistics across channels; **2)** the selected trials and channels were individually inspected and evaluated; if a trial or a channel was judged too noisy, it was definitively rejected; otherwise it was included in the dataset for further analysis. **3)** The dataset was globally visually inspected again to check for noisy channels or trials that were not visible in the first phase.

Time-frequency spectrum estimation

Oscillatory power for frequency over time was estimated using Fourier transformations in sliding time windows with a fixed length of 500 ms that moved in steps of 50 milliseconds. Frequencies were selected from 2 to 40 Hz in steps of 2 Hz. To control for spectral leakage and apply frequency smoothing, we used a multi-taper algorithm, but we used only one taper (Hanning taper).

Multivariate analysis

We analysed the data using a multivariate approach by employing a support vector machine algorithm (SVM; see Pereira et al., 2009). In the neuroimaging field, SVM is a data decoding technique that performs discrimination between neural patterns (e.g. groups of voxels or feature arrays) generated by experimental manipulation (e.g. neural patterns associated with condition A and neural patterns associated with condition B). The algorithm assigns numerical values to the patterns (dissimilar patterns have dissimilar numbers) and finds the optimal hyperplane that separates the two classes (condition A and B) of interest. The peculiarity of SVM is that it takes into account those patterns that are closer to the hyperplane more than those that are far away from it. One can select groups of voxels of interest (ROI approach) or scan the entire brain to look for pattern discriminability. The latter is known as the searchlight method (Kriegeskorte et al., 2006). Multivariate methods have been mainly used for fMRI studies (Haxby, 2012), and only recently have been started to be applied to MEG data (King & Dehaene, 2014). We performed multivariate analysis using CoSMoMVPA (Oosterhof et al., in preparation), an MVPA toolbox for Matlab that uses LIBSVM (Chang & Lin, 2011) that provides a searchlight method adapted for MEG data. The main difference from the fMRI approach is feature selection. In the case of MEG, a

neural pattern is the frequency-band intensity of a certain oscillatory activity over time and the sensors involved. Specifically, in the case of time-frequency representations, a neural pattern is a triplet of sensor-frequency-time power with its respective neighbours (for a more detailed explanation of the algorithm, refer to Chapter 2). Once the features have been selected, the SVM algorithm tries to differentiate the patterns associated with the conditions of interest.

We decided to use multivariate methods for two main reasons: **1)** they have been recognized to be more sensitive in discriminating between stimuli of the same class (e.g. distinguishing between two types of actions); **2)** they allow investigation at different levels of abstraction by means of cross-condition methods. For example, one can discriminate between actions performed with different effectors, meaning between two actions performed with the hands and two actions performed with the feet (within-effector decoding); or one might be interested in a higher level of abstraction, in representations of actions that are independent of the way in which they are performed (across-effector decoding).

We aimed at targeting neural representations with different levels of abstraction and thus used the cross-condition neural decoding method. Specifically, we designed the experiment for distinguishing between an abstract level (encoding actions irrespective of the way they are performed) and a more concrete level (encoding actions within a specific effector). To decode abstract actions, a classifier was trained on discriminating between two actions performed with one effector (e.g. discriminating between squeezing and throwing) and tested on discriminating between the two actions performed with the other effector (e.g. discriminating between crushing and kicking). To decode concrete actions, a classifier was trained to discriminating between two actions performed with one effector and tested at discriminating the same two actions performed with the same effector.

Statistical significance was assessed using a permutation test, a nonparametric method based on Monte Carlo simulation and a cluster-method to control for multiple

comparisons (Maris & Oostenveld, 2007). A more detailed description of the method can be found in Chapters 2

To evaluate the sources that accounted for any observed decoding effect at sensor level, we also run multivariate methods at source level. Frequency and time of interest were selected based on univariate contrasts between actions and body parts (torsos). Source power at each voxel and for each trial was estimated using the dynamic images of coherent sources, or DICS, (Gross et al., 2001) a beamformer method that looks for linear combinations of weights (spatial filters) that best explain the observed data at the sensor level. Thus, the resulting data resembled fMRI data and we could apply a searchlight approach.

Results

Effector-dependent representations of observed actions

To look for neural oscillatory patterns that encoded information related to observed actions performed with a specific effector, we performed a within-effector multivariate analysis (see Materials and Methods). We conducted two separated analyses for foot and hand actions. In both cases, significant decoding accuracy was observed in low frequency bands (between around 2 to 12 Hz) in a time period that immediately followed the stimulus onset and was mainly localized in posterior sensors (see **Figure 3.4**). We evaluated the significance of the classification accuracy by a permutation test (see Materials and Methods). The statistical analysis confirmed significant decoding in low frequency bands in the early time window (**Figure 3.4**, panel **C** and **G)**. Furthermore, significant differences were evident over time also at slightly higher frequency bands (between around 12 Hz and 18 Hz; see **Figure 3.4**, panel **D** and **H**), when posterior and frontal sensors appeared to show a preference for actions performed with the foot, whereas central sensors preferred actions performed with the hand. Topographical representations over time for the different frequency bands (see **Figure 3.5**, panel A for within-foot decoding and panel B for within-hand

decoding) helped at identifying which sensors were most sensitive for these two tests and highlighted bilateral decoding for actions performed with the foot, and a slight bias to posterior left sensors for actions performed with the hand. As is well known in the MEG community, topographical maps do not necessarily reflect the location of the underpinning neural sources because all sensors are somewhat modulated by the active neuronal population. Furthermore, from the topographical representation it is impossible to tell if the decoding effect was due to activity coming from parietal, occipital, or temporal regions, or even from a combination of all these regions. For this reason, we ran a multivariate analysis also at source level. Results for the earlier time window and the low frequency band are shown in **Figure 3.8** (Panel A and B) and revealed a bilateral involvement of occipito-temporal regions in generating low frequency effects in the early time window in both tests. Sources accounting for the decoding effect observed at later time points at higher frequency bands (beta) were localized in parietal and frontal regions **Figure 3.8** (panel D and E). Specifically, the within-foot test revealed sources located close to the hemispheric midline of the left and right central sulci, together with left prefrontal and right occipito-temporal cortex; the within-hand test showed the involvement of the left lateral and ventral sections of the central sulcus, along with right dorsal frontal cortex.

Effector-independent representations of observed actions

To evaluate the frequency bands that carried information related to effector-independent action representations, we ran an across-effector multivariate analysis (see Materials and Methods). The accuracy maps (**Figure 3.6**) showed similarities with the within-effector analysis, since posterior sensors showed good classification accuracy at low frequency bands immediately after stimulus onset. Interestingly, the decoding accuracy seemed more condensed over the posterior sensors. The permutation test confirmed these observations (corrected $p<0.05$), whereas no significant classification was observed at later times and/or at higher frequency bands. In addition, the topographical representations pointed to a bilateral preference for this type of representation (**Figure 3.7**). The source analysis again revealed bilateral sources at lateral occipito-temporal cortices accounting for the low frequency band

classification effect observed at sensor level (**Figure 3.8C, F**). Sources at later time points were localized in left inferior parietal and right temporal regions.

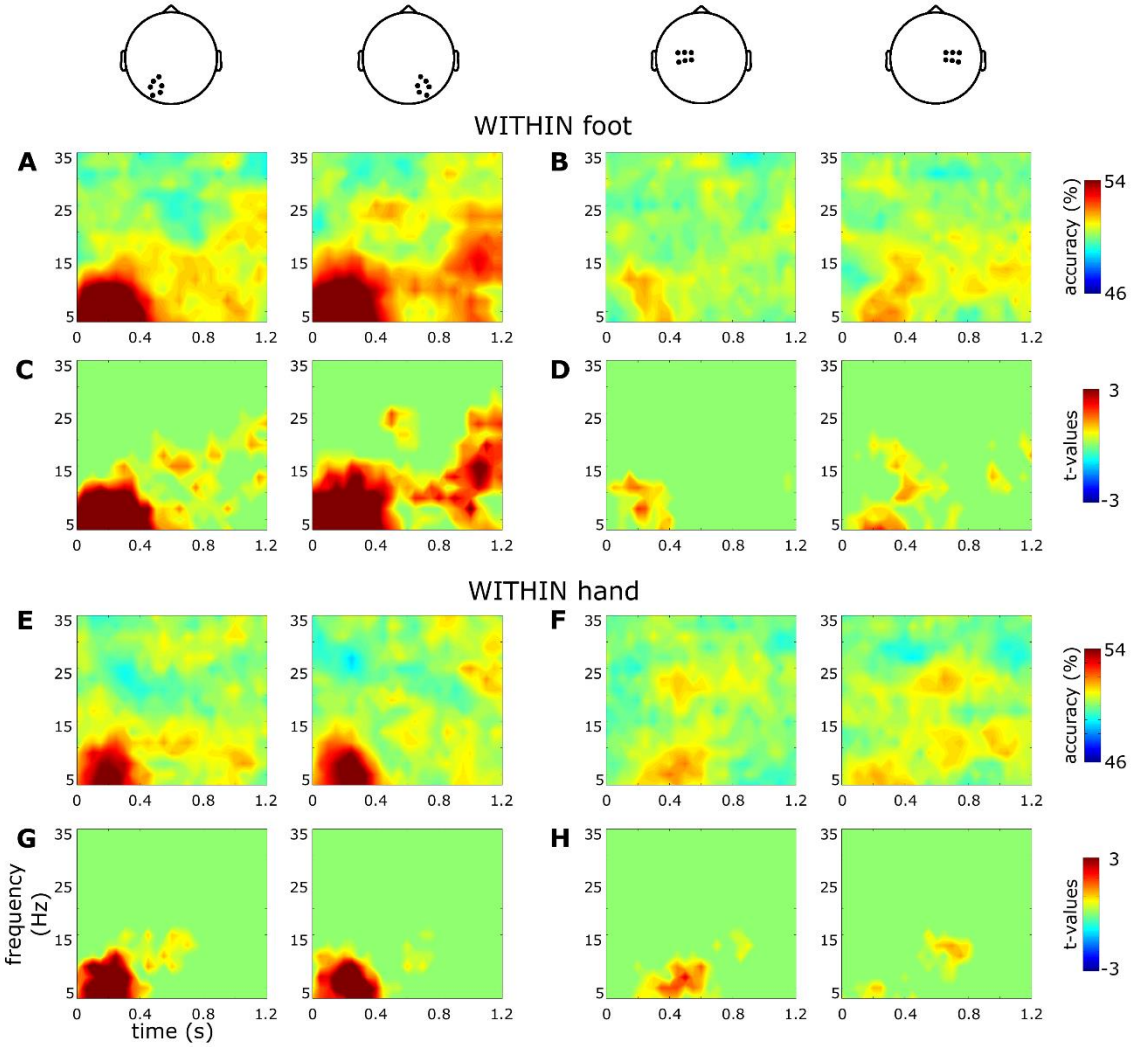


Figure 3.4. Accuracy and t-maps of the within-effector analyses. Time-frequency representations for the mean accuracy maps, and related masked t values, obtained from the within-foot analysis averaged from four groups of representative posterior (panel **A** and **C**) and frontal (panel **B** and **D**) sensors (see topoplots in the upper part of the figure). Similar plots were obtained for the within-hand analysis (posterior sensors: panel **E** and **G**; frontal sensors: panel **F** and **H**). To determine significant decoding (Panel **C**, **D**, **G**, and H), we ran a cluster-based Monte Carlo permutation test with 1000 iterations (Maris & Oostenveld, 2007). Zero indicates stimulus onset.

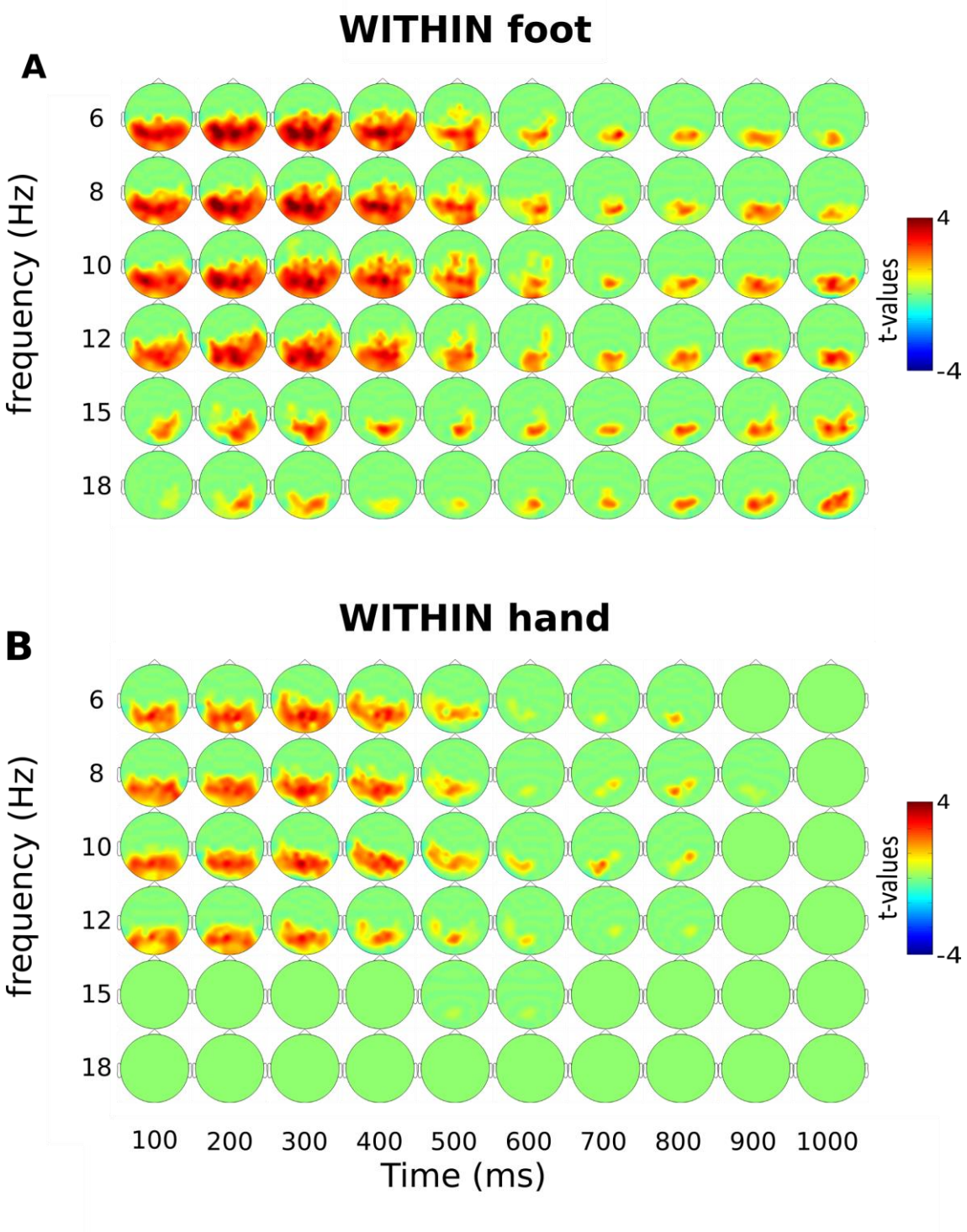


Figure 3.5. Neural dynamics of within-effector decoding. Topographic representations of classifier accuracy to discriminate between actions performed with the foot (panel **A**) or the hand (panel **B**) expressed as t values. The classifier was trained to discriminate within effector actions and tested on an independent dataset using actions performed with the same effector. Statistical tests (p-corrected < 0.05) were performed using a cluster size Monte Carlo permutation test (Maris & Oostenveld, 2007). Values that did not survive the multiple comparison correction are masked.

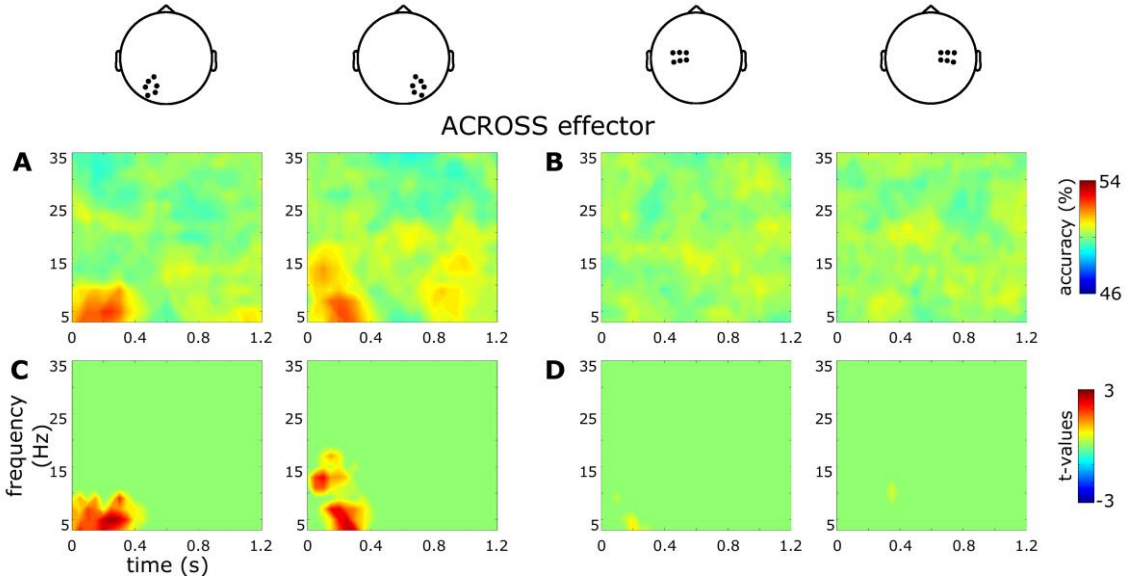


Figure 3.6. Accuracy and t-maps for the across-effector analysis. Time-frequency representations for the across-effector analysis which involved training and testing a classifier to discriminate between semantically related actions (*compressing an object/moving something away from the body*) performed with different effectors. Panel **A** and **B** show decoding accuracy in representative posterior and frontal sensors, respectively. Decoding accuracy at chance is 50%. Panels **C** and **D** show the corresponding t maps for t values that survived a Monte Carlo permutation test with 1000 iterations (Maris & Oostenveld, 2007).

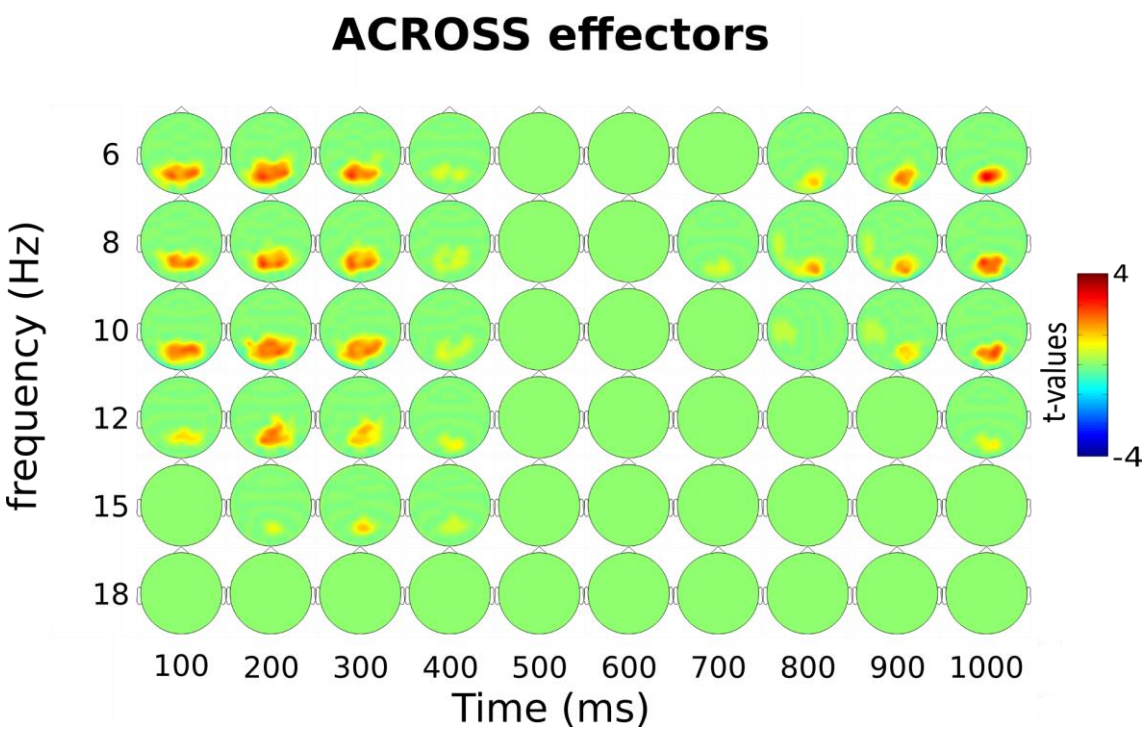


Figure 3.7. Neural dynamics of across-effector decoding. Topographic representations of classifier accuracy to discriminate between actions performed with across effector expressed as t values. The classifier was trained to discriminate actions performed with one effector (e.g. foot) and tested on actions performed with the other effector. Statistical tests (p-corrected < 0.05) were performed using a cluster size Monte Carlo permutation test (Maris & Oostenveld, 2007). Values that did not survive multiple comparison correction are masked.

Frequencies and sources overlap.

To get a clearer idea of the topographical differences between the three different tests, we overlapped the three significant accuracy maps on a sensor layout (**Figure 3.9**). The three tests mostly overlapped (orange colour in the image) in the posterior sensors at low frequencies and during the early time window, as expected. There was a preference for posterior and frontal sensors also for the within-foot test (green), especially at a late time window and a slightly higher frequency. Aside from posterior sensors, there were central and left frontal sensors at low and higher frequencies showing significant decoding for the within hand test (blue). The red colour, indicating across-effector decoding, is not very visible because it is almost entirely part of the overlap (orange colour) in the posterior sensors at low frequencies and the early

latency. These results suggest that the neural representations that these three tests are targeting are at least partially overlapping. The three source analyses we conducted for the three tests support this hypothesis given the similarity between the source representations observed in bilateral occipito-temporal regions. Such similarity was further exploited when we examined the overlap of the three source maps, which revealed common neural representations in bilateral occipito-temporal regions (**Figure 3.10**).

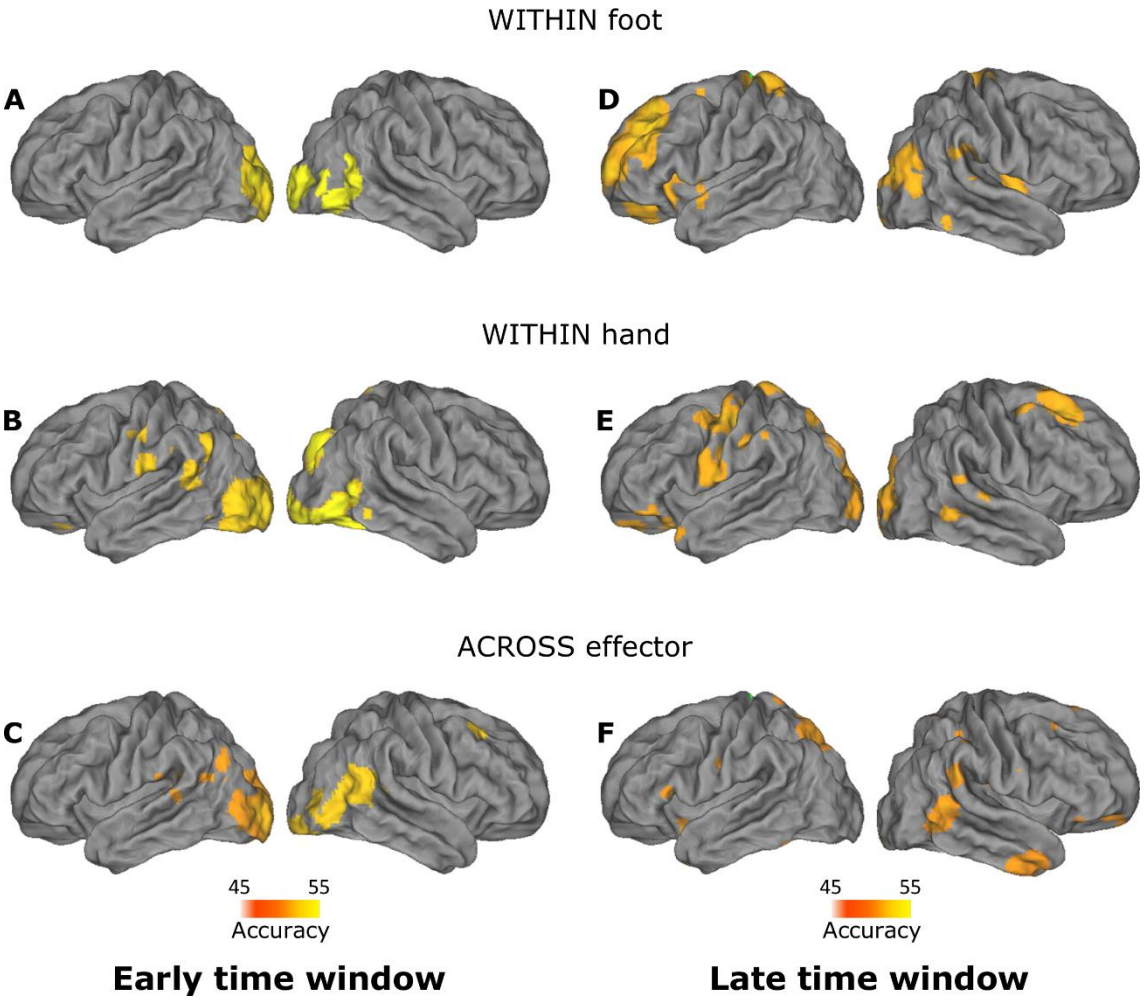


Figure 3.8. Accuracy maps in source space. To identify the sources that account for the decoding effects observed in sensor space, we ran multivariate analyses for the three tests in source space. We distinguished between an early (-50 to 450 milliseconds) and a late (474 to 825 milliseconds) time window. Based on the sensor results, we selected a low frequency band (centred around 6Hz, smoothed with 3Hz) for the early time window and a higher frequency-band (centred around 17Hz, smoothed with 3Hz) for the late time window. Only the top 10% accuracy values are depicted for visualization purposes. Maps were projected on the PALS atlas (Van Essen, 2005), using Caret software (Van Essen et al., 2001).

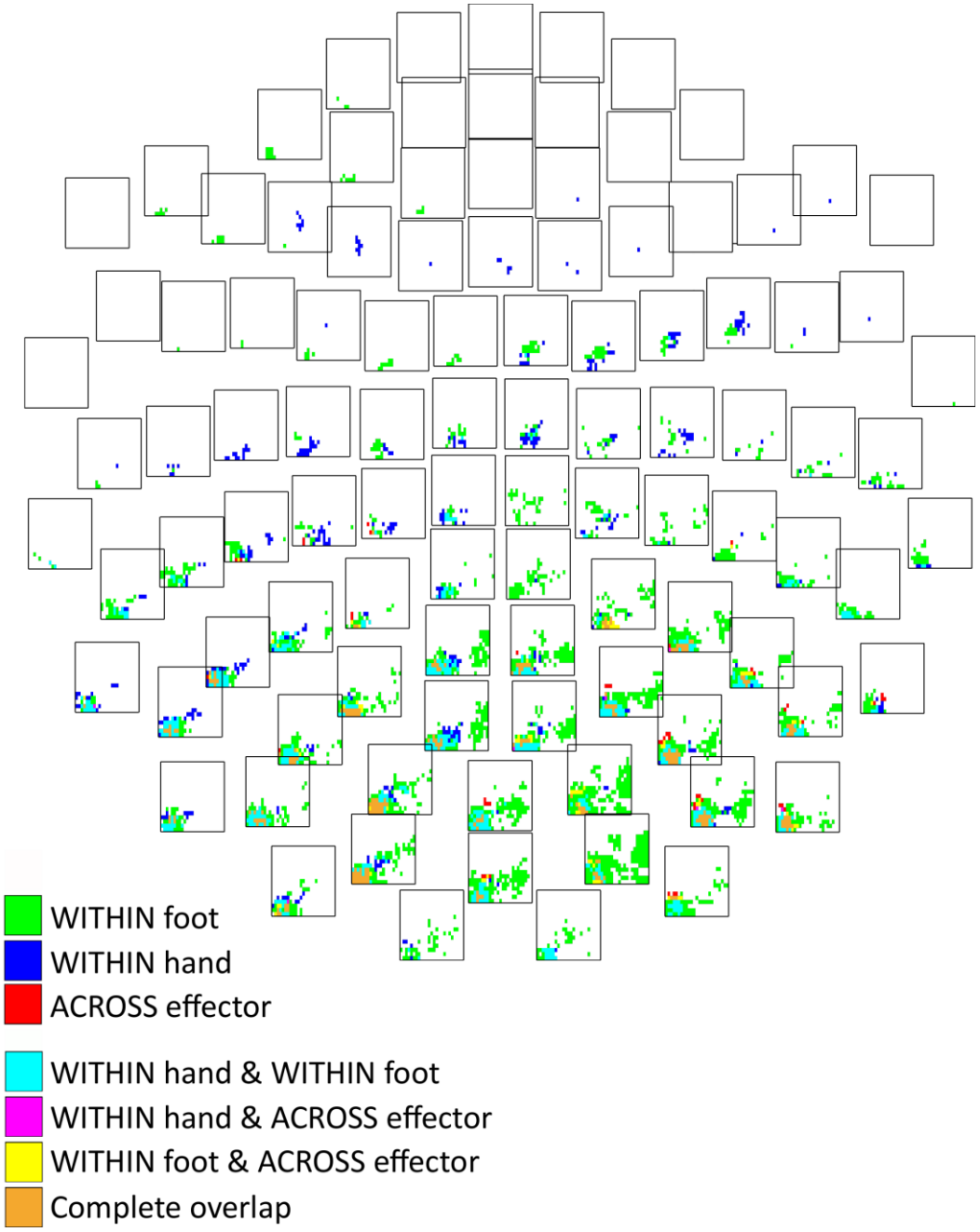


Figure 3.9. Accuracy overlap between tests in sensor space. To show similarities and differences between the three tests, we overlaid the three masked t maps (see **Figure 3.4C, D** and **Figure 3.6G, H**). Complete overlap between the three t maps (orange color) was observed in bilateral poterior sensors in the early time window. Main differences were observed between the within-hand (blue colour) and the within-foot (green colour) analyses over frontal sensors at later latencies. The across-effector analysis was confined to posterior sensors.

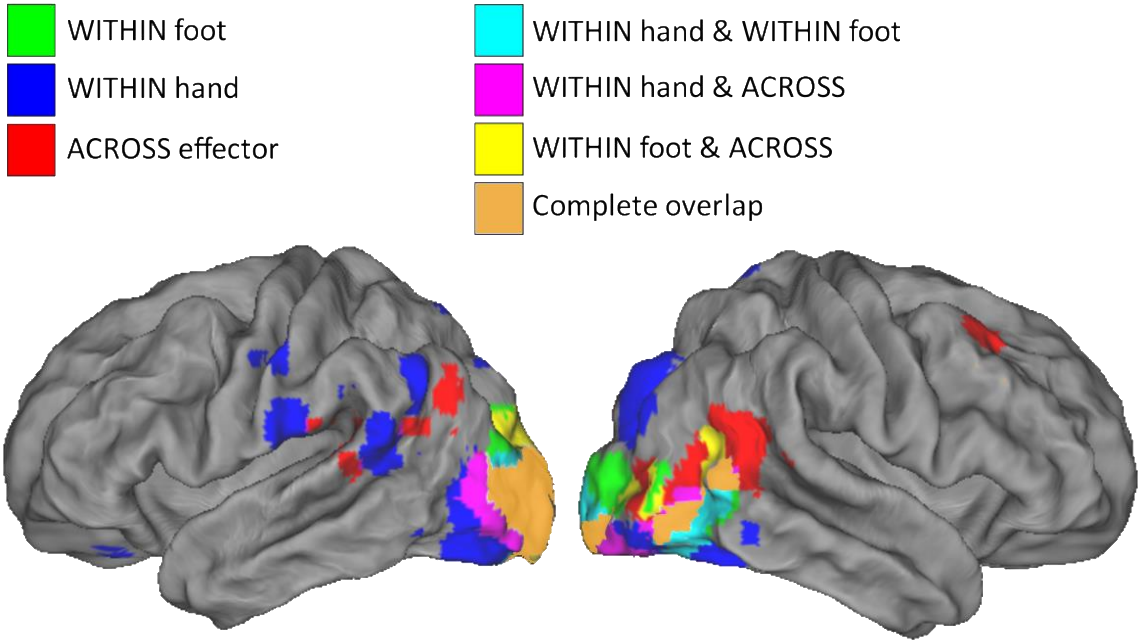


Figure 3.10. Overlap between the three tests in source space. The figure shows the overlap between the three accuracy maps obtained from the multivariate analysis performed in source space (**Figure 3.8**) to account for significant decoding effects observed at sensor level (**Figure 3.4** for the within-effector analyses, and **Figure 3.6** for the between effector analysis). A complete overlap between the three types of analysis was observed within lateral occipito-temporal regions. The maps are masked based on the top 10% accuracy values.

Discussion

In this study (**Study 2**), we investigated the neural dynamics of action representations during observation of hand-related and foot-related actions. The main aim of this study was to follow up on important findings observed in **Study 1** (chapter 2). In **Study 1**, we observed that temporal regions, specifically LOTC, and parietal regions, most likely IPL/IPS, have an earlier access to abstract action representations than frontal motor-related regions. However, in **Study 1**, we only used two actions (grasping/pointing) and decoding generalization was across movement direction and hand used only, but a region involved in action recognition should be able to discriminate between a richer set of actions at higher abstract levels

Therefore, in **Study 2,** we used four new actions, two of which performed with the hand (squeezing/throwing) and the other two with the foot (crushing/kicking). These actions can be distinguished within effector (squeezing vs throwing/ crushing vs kicking) to investigate a concrete action representation (effector level). Furthermore, one can discriminate these actions across effectors (*squeezing vs crushing*/*throwing vs kicking*) as they share similar concepts (*compressing an object*/*moving something away*). In this way, one can target a more abstract level of representation. To this aim, we used MEG in combination with multivariate analysis to investigate the neural dynamics of abstract and concrete action representations. We observed abstract (effector-independent) and concrete (effector-dependent) action representations in LOTC in a time window immediately subsequent to stimulus onset; we also observed concrete (effector-dependent) representations only in frontal effector-specific motor-related areas in a later time window.

These results corroborate and extend the findings of **Study 1**: they confirm that LOTC contain information related to abstract action representations and that this information is accessed earlier than frontal regions. Furthermore, encoding of action representations in frontal motor-related regions is modulated by the level of abstractness. Specifically, frontal regions appeared to encode for abstract action representations in the sense that they showed generalization across reach direction, and from one hand to the other. By contrast, the current study indicates that no such generalization is observed in frontal regions from hand to foot and vice versa. These findings are partially in line with two recent fMRI-MVPA studies that showed that frontal motor-related regions encode concrete action representations (Wurm & Lingnau, 2015; Wurm et al., 2015), whereas LOTC hosts abstract action representations, most likely because this region contains action concepts. However, the representation we observed in motor-related areas in our previous MEG experiment was abstract as it generalized across hand used and movement direction. What could explain the inconsistency between these fMRI experiments and the results of **Study 1**? There are two potential explanations. Following Wurm et al. (2015) and Wurm & Lingnau (2015), the way we addressed abstract action representations was

different from that of Wurm et al. (2015): instead of generalizing across the object on which an action was performed, **Study 1** aimed to generalize across reach direction and effector (left vs right hand). The second explanation is related to the types of actions that we used. Grasping and pointing are very common actions, since we are exposed to these actions for most of our daily life. This means, following the association hypothesis of mirror neurons (Heyes, 2010), that these two actions are over-represented in the semantic and motor system and low-level processing might be triggered more automatically.

Chapter 4: Open object = Opening (Closed object)

Adapted from:

Tucciarelli, R., Wurm, M., Roth, Z., Zohary, E., and Lingnau, A. (in preparation). *Dissociating action states and functions – an MVPA study.*

Introduction

Think about the example provided in the general introduction: You are watching a friend making coffee. Which are the most salient features in the scene? Focus just on the initial part of the preparation, grasping and opening the moka. There are at least two important aspects here: the two actions (grasping and opening) and the object. If you think deeply, you will see more: the configuration of the object changes after the actions have been performed. 'Grasping', will bring the moka to a new position; after 'opening', the moka will be not closed anymore.

Generally speaking, we can define an object-related action as a function (**Figure 4.1**) that changes the state of an object. For example, opening a moka pot changes the object state from closed to open. In mathematical terms, an action is a relationship between a set of inputs (initial object states) and a set of possible outputs (end object states):

$$f: X \rightarrow Y \text{ or } y = f(x)$$

$$open\ object = opening\ (closed\ object)$$

Open object = Opening (Closed object)

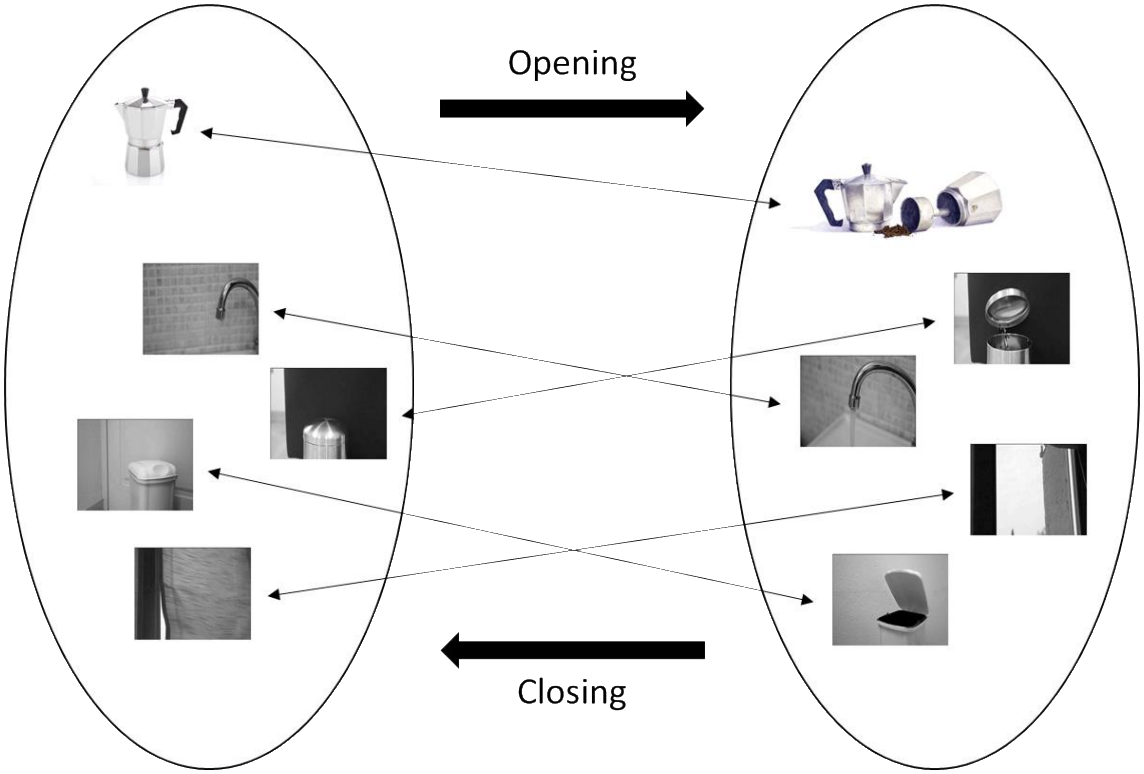


Figure 4.1. Actions as functions that associate object states. An action can be described as a *function* that changes the state of an object. For example, the action *opening* will change the state of a moka moka pot from *closed* to *open*; the function *closing* will change the state of an *open* trash bin to *close*.

Studies 1 (Chapter 2) and **2** (Chapter 3) showed that LOTC and IPL encode abstract action information. Specifically, neuromagnetic oscillatory activity generated within temporal and parietal regions contains information related to an observed action that generalizes across action movements and effectors (right, left hand, and foot). Notably, the high temporal resolution of MEG allowed us to show that such action representations are encoded in temporal regions earlier than in frontal regions, where concrete action representations are represented (see also Kilner, 2011 for a functional explanation of concrete encoding within motor-related regions). These findings suggest that LOTC and IPL play an important role in action concept encoding, and this view is supported by other fMRI (Binder et al., 2009; Fairhall & Caramazza, 2013; Hickok, 2012; Lingnau & Downing, 2015; Wurm & Lingnau, 2015; Wurm et al., 2015) and patient (Kalénine et al., 2010; Tarhan et al., 2015) studies. To date, however, it is

unclear whether LOTC and IPL encode the function underlying actions (e.g. opening something), the end state (e.g. the opened object), or both. In an fMRI adaptation study, Hamilton and Grafton (2008) presented participants with videos of actions (e.g. opening a sliding box) with same or different outcomes that used the same or different kinematics. They found that frontal (IFG) and parietal (aIPS) regions were sensitive to the outcomes of an action, with similar trends in the LOTC (p=0.053). Critically, both the action and the final state of the object were always presented in the same scene, making it impossible to disambiguate between object and action information at the neural level.

The goal of **Study 3** was thus to examine whether LOTC and IPL preferentially encode functions or object states. In two separate fMRI experiments, we used specific object categories (trash bins/window blinds/taps) of objects that allowed us to present different aspects of the scene: the object state only (*Experiment 1*) and the action only or both the action and the object state (*Experiment 2*).

In *Experiment 1*, we investigated what brain regions contain information related to object states (open/close). Specifically, we were interested in abstract object state representations, that is, representations that generalize across specific objects or to low-level object features (e.g. perspective). In *Experiment 2*, we examined the representation of action functions with the object state either visible or not visible. This allowed us to distinguish the two types of representations at the neural level and individuate eventual overlaps. To this aim, we adopted a data-driven searchlight approach.

We looked for areas where representations of action functions and action states either 1) were dissociated at the neural level (no overlap); or 2) shared common neural substrates (partial or total overlap). These two possible outcomes directly address predictions made by the two theories of action understanding: the motor theory predicts motor regions will be involved in action functions and states, whereas the cognitive theory predicts this information will be encoded in temporal regions (Hickok,

Open object = Opening (Closed object)

2012; Lingnau & Downing, 2015) and/or parietal (Binder et al., 2009; Fairhall & Caramazza, 2013).

We found no evidence that the ventral premotor cortex (PMv) encoded either abstract action functions or action states. LOTC encoded both action functions and object states, while IPL encoded action functions only. These findings provide important information regarding the organization of action semantics in the brain.

Materials and Methods

Participants

Twenty-one participants took part in the experiment. All participants were right-handed with normal or corrected-to-normal vision. None of them had history of neurological or psychiatric disorders. Participants gave their formal consent to participate to the study, which was approved by the Helsinki Ethics Committee of Hadassah Hospital, Jerusalem, Israel.

Stimuli

The stimulus set (see Appendix for the complete list) consisted of still images (Experiment 1) and videos (Experiment 2). We first filmed the actions in two conditions: either the entire scene (*object state visible*) or the action alone (*object state invisible*). The videos started in the moment in which the effector was already touching the object, just some millisecond before movement initiation and ended some millisecond after the movement finished. Therefore, a typical video with the object state visible contained both the initial (frame 1) and the final (frame 30) object states. We then extracted these two frames to create the images for Experiment 1. The frame was cropped in a way that only the critical part of the object showing the state was visible. The images depicted the three object categories (window blind/trash bin/and tap) in the two possible states (open/close). For each object category, there were three different exemplars in two different perspectives. Each specific image/video was repeated 18 times. Therefore, each object state was repeated 18 x

Open object = Opening (Closed object)

3(objects) x 2(perspectives) = 108 times. The videos contained actions related to the objects used in Experiment 1. Two actions were used (opening/closing). The actions were filmed from two different perspectives. The total number of videos was also 108. Images and videos were captured using a Canon 5D Mark II camera and edited in iMovie (Apple) and MATLAB (The MathWorks). Stimuli were in black and white, had a length of 2s (30 frames per second), and had a resolution of 400x300 pixels.

To disentangle action functions from object states, we selected a set of items that allowed us to show these two aspects separately. We thus used objects whose state could be changed without directly manipulating the critical part that indicates whether it is opened or closed. For example, a trash bin can be opened or closed with a hand by directly acting on the lid, or one can open the lid by pressing the handle with the foot. Critically, the handle is relatively far from the lid. In Experiment 1, we thus showed only the part of the object defining its state. In Experiment 2, the videos were presented in two critical conditions: in one case, both the object state and the action were visible (state-visible condition); in the second case, only the action but not the object state was visible (state-invisible condition), as the field of view was restricted to the action. Since we aimed to target at abstract action and object state representations, we selected stimuli with high variability of low-level features (different objects/different perspectives/different location/different actors) such that the observed results cannot be explained as classification of low-level visual properties of the stimuli. Visual stimuli were projected via an MR-compatible projector onto a screen located 114 cm from the participants. The screen was made visible to the subjects via a tilted mirror, positioned above the subjects' faces. The screen was 53 x 30 cm (26.4° × 15.1°). The display resolution was 1920 × 1080 pixels. Stimulus presentation was controlled using ASF (Schwarzbach, 2011) a Matlab toolbox based on Psychtoolbox (Brainard, 1997).

Experimental task

To ensure participant's alertness, participants had to detect the presence of ambiguous images or videos (catch trials). In Experiment 1, catch trials consisted of

images depicting objects that could not be classified as being in either an open or a close state such as a trash bin with the lid only slightly raised. Similarly, in Experiment 2, catch trials were videos showing actions that could not be classified as neither opening or closing. Before entering into the scanner, participants familiarized with the images of Experiment 1. They observed all the stimuli at least one time and had to indicate by button press the state of the objects and the catch trials. During the anatomical scan (see Experimental design), participants observed the videos used in Experiment 2 and had to indicate the type of actions and the catch trials.

Experimental design

We adopted an event-related design with null-trials to guarantee optimal statistical power (Dale, 1999) using optseq2 (https://surfer.nmr.mgh.harvard.edu/optseq/). The proportion of null-trials was computed as the total duration of the experiment divided by the number of the conditions (experimental conditions and catch trials). The mean number of null trials per run was 38.11 (SE: 4.49). The order of trials was pseudo-randomized. Each experimental trial consisted of a 2 second stimulus (either an image or a video) followed by a fixation cross (1 second). Null trial duration varied from 1 to 6 seconds. The two experiments were conducted within one scanning session (Figure 4.2). The session included nine functional runs and one anatomical run organized as follows: three functional runs (Experiment 1) + one anatomical run + six functional runs (Experiment 2). Five subjects performed only four functional runs in Experiment 2. The order of the experiments was kept the same for all participants to ensure that they did not activate previously seen action representation.

Open object = Opening (Closed object)

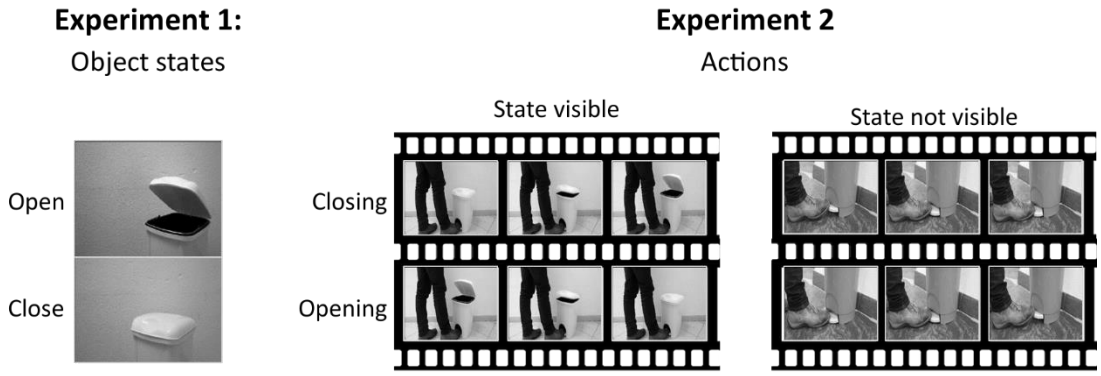


Figure 4.2. Experimental design. In Experiment 1, we presented images of objects in two possible states (open/closed). In Experiment 2, participants watched video clips of actions performed with the objects presented in Experiment 1. In the videos, the object state was either visible (visible-state condition) or not visible (invisible-state condition). Note that the object states (images) were taken from the video clips (initial and final frames) after cropping the image. Participants had to attentively observe the stimuli and report catch trials, namely objects which were in neither an open nor a closed state (Experiment 1), or actions that were neither opening nor closing (Experiment 2).

fMRI data acquisition

The blood oxygen level-dependent (BOLD) fMRI measurements were obtained using a whole-body 3-T Magnetom Trio Siemens scanner and a 32-channel head coil. The functional MRI protocols were based on a multislice gradient-echo planar imaging and obtained under the following parameters: time repetition (TR) = 2 s, time echo = 30ms, flip angle = 90°, imaging matrix = 64 × 64, field of view= 192mm; 37 slices with 3-mm slice thickness and 0.35mm and were oriented in the oblique position, covering the whole brain, with functional voxels of 3 × 3 × 3 mm. In addition, high-resolution, T1-weighted magnetization-prepared rapid acquisition gradient-echo (MPRAGE) images were acquired (1 × 1 × 1 mm resolution).

Preprocessing

Functional data were preprocessed using BrainVoyager 2.8 (BrainInnovation) in combination with Neuroelf (http://neuroelf.net/) and custom scripts. The first four volumes were discarded online to avoid T1 saturation. Data were 3D motion corrected

using the first retained functional volume as reference. Functional data were high-pass filtered (cut-off frequency of three cycles per run) and slice time corrected. Functional data were smoothed using a 8mm Gaussian kernel for univariate analysis only. The functional volumes were then aligned to the anatomy after Talairach transformation using trilinear interpolation.

Univariate analysis

Classical univariate analyses were conducted separately for the two experiments to localize regions modulated during the observation of images (Experiment 1; functional contrast: all images vs baseline) and videos (Experiment 2; functional contrast: all videos vs baseline). We computed a group random-effects general linear model (GLM) using a design matrix with 3 predictors of interest (action: open/close and catch trials) for Experiment 1, and 5 predictors (action x object state: visible-state opening/visible state closing/invisible-state opening/invisible-state closing and the catch trial) for Experiment 2. The two design matrices also contained the 6 motion parameters. Each predictor was convolved using a hemodynamic model (dual-gamma). Each trial was 2 second long. This time course model was used to fit the fMRI signal for each voxel. We thus obtained a β-value for each voxel that were used to compute the functional contrasts. Results were corrected for multiple comparisons using the Bonferroni procedure.

MVPA

We conducted a multivariate volume-based searchlight analysis using a support vector machine (SVM) classifier using the LIBSVM (Chang & Lin, 2011) implementation embedded in CoSMoMVPA (Oosterhof et al., in preparation). We performed the multivariate analysis to individuate those regions that encode abstract action states (Experiment 1) and abstract action functions (Experiment 2). To this aim, we trained a classifier to discriminate open and close state (Experiment 1) or opening and closing action (Experiment 2) from a subgroup of stimuli (Figure 4.3 shows a schematic example of MVPA for Experiment 1). Then we tested the accuracy of the classifier in categorizing an independent subset of stimuli. The predictors used as input for the

multivariate analysis were estimated on the basis of three trials in a GLM analysis. In Experiment 1, for each of the seven runs we had six predictors (β-weights) per condition per run. In Experiment 2, there were three predictors per condition per run. The analysis was restricted to grey matter and the searchlight radius was 4 voxels (mean size: 219.5 ± 43.1 voxels). We adopted a *leave-one-run-out cross-validation* approach. All possible combination of training and testing patterns were conducted and the final accuracy was obtained averaging the accuracy maps resulted from the all iterations. The decoding analysis was performed at individual level and then the individual accuracy maps were averaged across subjects.

Open object = Opening (Closed object)

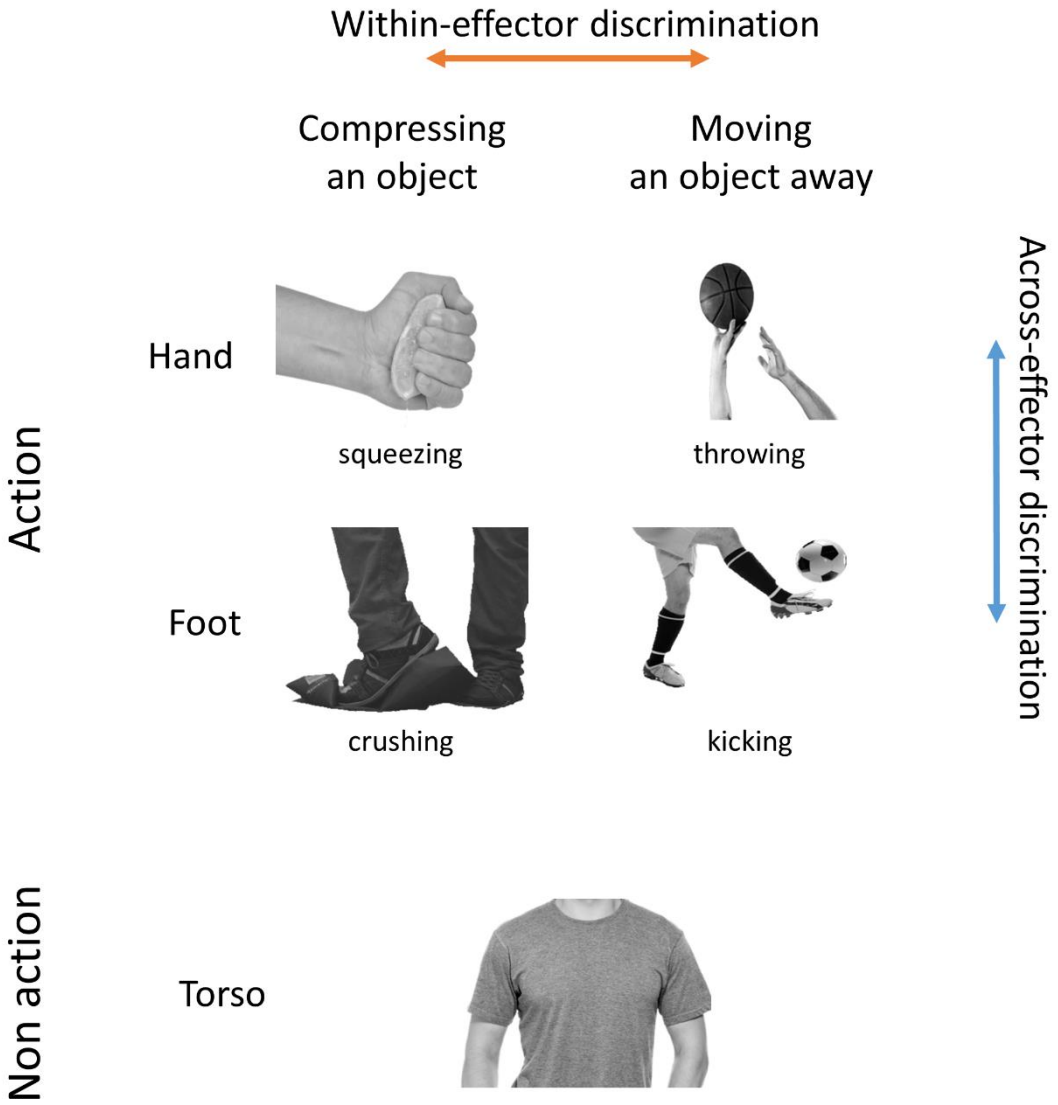


Figure 4.3. Schematic representation of MVPA. We looked for brain regions that contained neural patterns related to object states (Experiment 1) and action functions (Experiment 2) by running a volume-based searchlight (Kriegeskorte et al., 2006) multivariate pattern analysis (MVPA) using a support vector machine (SVM). The figure shows a hypothetical pattern elicited within temporal regions when observing object states (the example holds for video clips). If a region contains information about a certain condition (e.g. object state), the elicited patterns should be similar when the object states are the same (e.g. open) and dissimilar when the object states are different. Analyses were conducted using CoSMoMVPA (Oosterhof et al., in preparation).

To assess statistical significance of the group accuracy maps, we run a one-sample t test, in which we compared the accuracies against chance level (50%). To account for

multiple comparisons, volume t maps were corrected at a cluster level (p=0.05) based on Monte Carlo simulation (1000 iterations). The initial voxelwise threshold was p<0.005. Accuracy and t volume maps were transformed to surface mesh vertex for visualization purpose.

To show overlap between accuracy maps obtained in the different tests, a cluster-level test was run at the surface level for each map (initial threshold p<0.05). All significant vertices were assigned with a constant number and a different color.

Results

Behavioural results

Sixteenth of the 21 participants (we could not collect behavioural responses of the first five subjects) identified the catch trials with high accuracy. The mean accuracy in Experiment 1 was 92.36% (SE: 1.4%), whereas in Experiment 2 it was 90.56% (SE: 1%). A pairwise t test did not show evidence for a difference between the accuracies of the two experiments (t(15); p-value=0.3659).

Univariate fMRI results

In both experiments, the univariate analyses revealed similar widespread activations within temporal, parietal and frontal regions, therefore classical regions of the action observation network (Caspers et al., 2010) were recruited. Specifically, for the images vs baseline contrast (**Figure 4.4**, panel A), we observed bilateral LOTC, bilateral IPS/IPL, left IFG and right LPFc. For the videos versus baseline contrast (**Figure 4.4**, panel B), we found bilateral LOTC, bilateral IPLS/IPL and bilateral IFG. In parietal and temporal regions, the two contrast maps were quite similar, as revealed by the overlap (**Figure 4.4**, panel C).

Open object = Opening (Closed object)

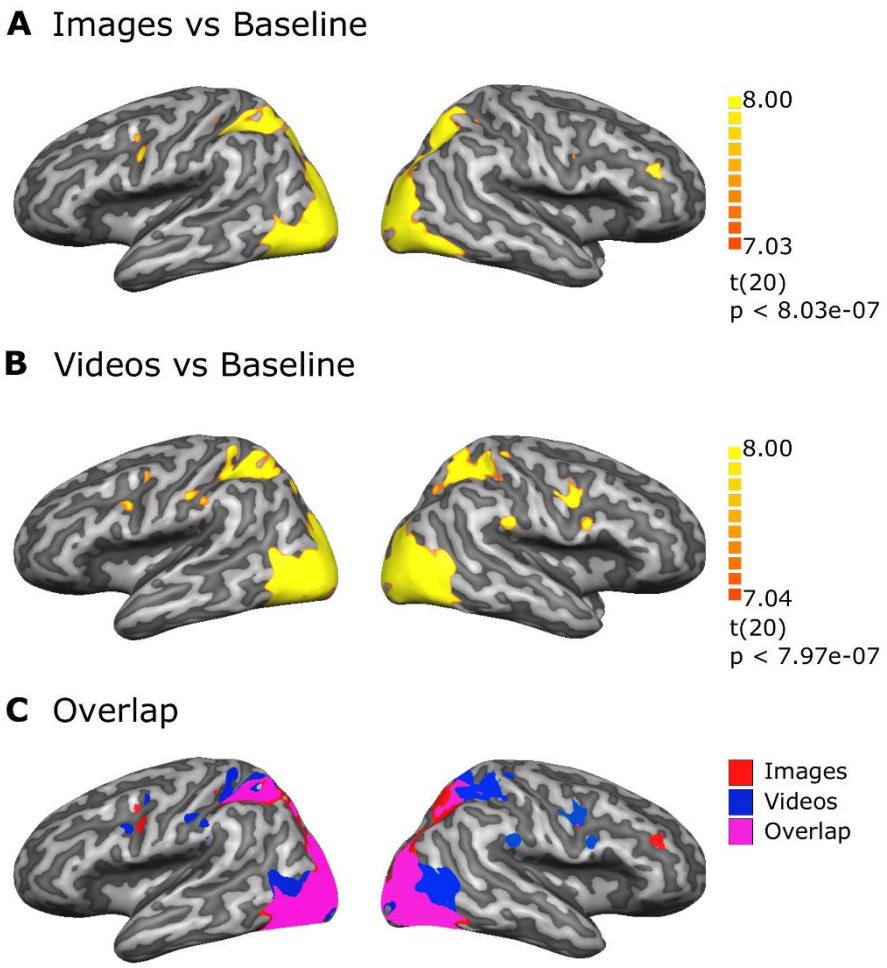


Figure 4.4. Univariate analysis. Univariate analyses were performed to locate regions recruited while viewing the images (Experiment 1) and the videos (Experiment 2). We computed a random-effects (RFX) GLM contrast of all images versus baseline (panel A) and all videos vs baseline (panel B). Both contrasts (t maps) revealed widespread activations in temporal, parietal and frontal regions, resembling the typical action observation network (Caspers et al., 2010). In particular, in both experiments we observed a recruitment of bilateral lateral occipitotemporal cortex (LOTC) and bilateral IPS/IPL (see overlap in panel C, violet colour). Videos appeared to recruit a slightly wider network, also including bilateral inferior frontal gyrus (IFG). Images also recruited left IFG and right lateral prefrontal cortex (LPFc). Results are corrected for multiple comparisons using the Bonferroni procedure.

Searchlight fMRI results

Using a data-driven volume searchlight analysis, we looked for regions encoding either object states of an object (Experiment 1) or actions defined as functions that change the object states (Experiment 2). For visualization purpose, accuracy maps for the

three tests and the corresponding (not corrected) t maps projected on the surfaces are shown in **Figure 4.5**. Information related to the significant (cluster corrected) clusters related to the three tests are reported in **Table 4.1**.

We found significant decoding accuracy in regions typically involved in action observation (see Univariate analysis). Specifically, a ventral bilateral portion of LOTC contained information related to object states (**Figure 4.5**A, D), whereas a more dorsal right section of LOTC and left IPL contained information relate to the actions when the object state was visible(**Figure 4.5**B, E). We could not find any significant encoding for actions when the state was not visible. We therefore decreased the threshold until p=0.02 and we found significant accuracy in right LOTC (**Figure 4.5**C, F).

To show common neural substrates between processing actions and object states, we overlap the t maps related to the state and the action test in the state-visible condition. Functional overlap was observed in left LOTC.

Region	x	y	z	Size	t max	t mean	Accuracy
States							
Left LingG/LOTC	-23	-81	5	437	5.842741	3.399395	54.63
Right Ins	44	3	10	72	4.766046	3.480899	56.48
Left CingG	-4	17	29	63	4.730336	3.466814	56.48
State-visible actions							
Left MTG/LOTC	-34	-70	24	172	6.392787	3.486930	59.23
Left STS/LOTC	-50	-58	31	35	5.583564	3.511877	57.49
Left AG	31	-62	30	106	4.721751	3.342583	54.29
State-invisible actions							
Right MTG/LOTC	43	-59	16	98	4.699485	2.771144	55.48

Table 4.1. Table of significant clusters (size in mm^3).Threshold p=0.005 for "Encoding States" and "Encoding Actions State Visible"; Threshold p=0.02 for "Encoding Action State Invisible". Corrected cluster threshold p=0.05. LingG, Lingual Gyrus; Ins, Insula; CingG, Cingulate Gyrus; MTG, Middle Temporal Gyrus; STS, Superior Temporal Gyrus; AG, Angular Gyrus;

Open object = Opening (Closed object)

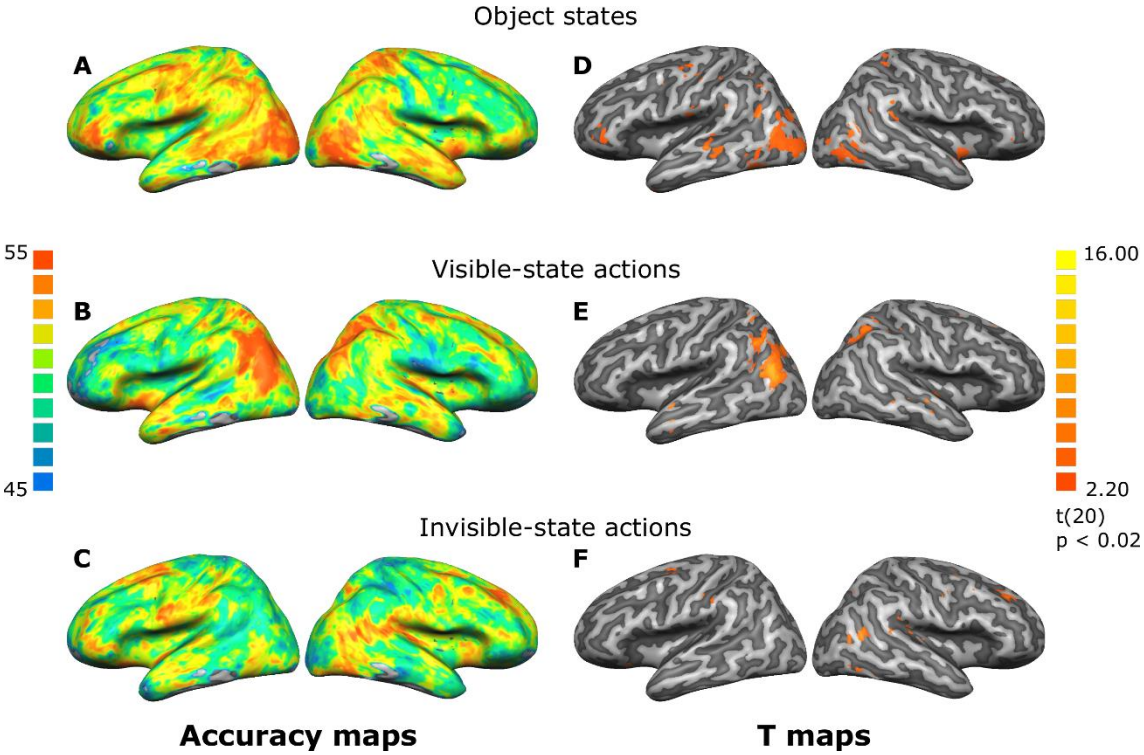


Figure 4.5. Accuracy and t maps for the three searchlight analyses. Mean accuracy maps (panels A, B, C) and related t maps (panels D, E, F) of the searchlight MVPA for the three tests of interest: discriminating object states (panel A and D; Experiment 1); discriminating actions when the object state is visible (panel B and E) and when the object state is not visible (panel C and F; Experiment 2). Chance accuracy level is 50%. For the t maps, threshold is p=0.02 (uncorrected). For corrected results refer to **Errore. L'origine iferimento non è stata trovata.**.

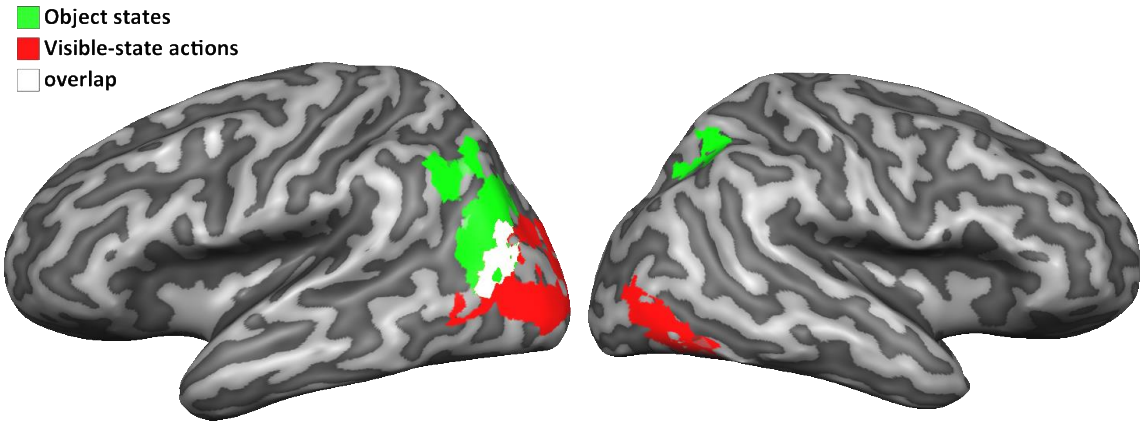


Figure 4.6. Overlap of the statistical maps. The statistical maps of the two tests of interests (decoding object states and decoding actions when the object state is visible) were overlaid on a surface brain. The reported clusters survived cluster size correction at initial threshold p=0.05 and corrected cluster threshold p=0.05.

Open object = Opening (Closed object)

Discussion

Action semantics refers to the knowledge related to actions, meaning "what the action is about, what its goal is, and how it is related to other objects" (Nelissen et al., 2005). An action can be defined as a *function* that changes the state of an object. To date, no previous study distinguished between two key features underlying the semantics of actions, namely, action functions and object states that typically co-occur. According to motor theories, mirror neurons subserve action semantics by using the observer's own motoric action repertoire to access the meaning of an action (Rizzolatti & Craighero, 2004). Therefore, encoding action functions should take place within motor-related areas. According to Nelissen's definition, object states could be encoded in the motor system as well, because observing an object in a particular state might recruit action knowledge "related to" the object. For example, observing a closed box could activate the action *opening*. In our study, we aimed to disentangle these two aspects by showing either the object states (Experiment 1) or the action functions with and without the object states (Experiment 2).

In a previous fMRI-adaptation study in which participants had to observe actions that had different object outcomes using the same or different kinematics, the inferior parietal lobule (IPL) and the inferior frontal gyrus (IFG) appeared to encode information related to action outcomes (Hamilton & Grafton, 2008). It is worth noting that LOTC showed a similar but non-significant trend. However, the authors used videos of actions in which both object changes and actions were present, making it impossible to disentangle between the two aspects.

In contrast to motor theory predictions, we did not find any evidence for encoding either action functions or object states in premotor regions. We observed that LOTC encodes information related to both action functions and object states, whereas IPL only contains representations of action functions.

In our study, action semantics did not appear to be encoded in premotor regions, but rather within temporal and parietal regions. Involvement of these areas in the encoding of action concepts has been reported in several previous fMRI (Oosterhof et

al., 2010, 2012a; Wurm & Lingnau, 2015; Wurm et al., 2015) and voxel-based lesion symptom mapping studies (Kalénine et al., 2010; Tarhan et al., 2015). Furthermore, a number of studies suggested a concrete to abstract gradient of representations running from posterior to anterior LOTC (Lingnau & Downing, 2015).

We were able to show that object states (open/closed) are encoded bilaterally in LOTC, whereas action functions with object states visible were encoded in left LOTC and bilateral IPL. We argue that IPL is most likely implicated in action function processing only, while left LOTC is involved in action semantics in general, irrespective of whether these semantics represent action-relevant object information or the function underlying the action.

We were not able to decode any type of information when comparing the two actions (opening/closing) in the invisible object state condition in left LOTC. This was a rather unexpected result, most likely related to task difficulty. Some participants at the end of the experiment reported difficulties to discriminate between the two actions in this condition, without also observing the initial and final object states. In other words, it is possible that we did not find any type of significant classification in this condition because participants were unable to distinguish between the two targets of interest (opening and closing). Note that we did find a significant decoding accuracy in right LOTC in this condition when using a more liberal threshold. Further experiments are necessary to examine whether these two types of conditions are actually encoded in different regions and if it is possible to decode action types without viewing the object state when the task is made easier also in LOTC.

Chapter 5: General Discussion

Which regions encode abstract action representations? Moreover, which are the neural dynamics of such representations? Using MVP analysis in combination with MEG and fMRI data, we found that temporal regions, specifically LOTC, encode abstract action information earlier than frontal regions. Furthermore, whereas LOTC showed generalization across reach direction, from left to right hand and vice versa, and from hand to foot and vice versa, frontal regions showed a late generalization across reach direction and from one hand to the other, but not across the type of effector. Finally, we found that LOTC encode both action functions and object states. Collectively, these results are hard to reconcile with a dominant role of motor-related frontal regions in action recognition. By contrast, our results indicate that LOTC might have a determinant role in action recognition.

Many neuroimaging studies have shown involvement of motor-related regions during action observation both in monkeys (di Pellegrino et al., 1992; Gallese et al., 1996; Umiltà et al., 2001) and in humans (e.g. Cattaneo et al., 2010; Fadiga et al., 1995; Muthukumaraswamy et al., 2004; Rizzolatti et al., 1996). These studies have also shown that some degree of abstraction exists in motor-related regions. For example, Gallese et al. (1996) reported monkey mirror neurons discharging for actions, such as *grasping*, independently of the way this action was performed (precision grip vs whole hand). In humans, a TMS study by Cattaneo et al. (2010) showed that PMv contains action information (pushing, pulling) that generalizes across the effector (hand, foot) used. Furthermore, Kilner et al (2009) reported a cross-modal (action execution to action observation and vice versa) adaptation effect in PMv. These results are seemingly in line with the claims made by motor theories that motor-related regions have a causal role in action recognition.

Our findings partially agree with these experimental evidences as we also observed motor-related regions being involved during action observation, but up to a certain degree of abstraction (e.g. across movement direction, but not across effector). We also extend those results as we added important information related to the neural dynamics of action abstract representations. We showed that abstract action information in motor-related frontal areas is provided only after action abstract encoding in temporal regions. These results conflict with motor theories which would have predicted early abstract action representations being *instantaneously* mapped within the motor system (Hauk et al., 2008).

Furthermore, we showed that the encoding of action information in motor regions might depend on the degree of abstraction that the analysis is targeting. For example, we were able to decode within-effector action information in motor-related regions, but not across-effector action information. These findings are in line with previous studies that used MVPA in combination with fMRI data, which found a generalization across viewpoint (Oosterhof et al., 2012a), tasks (Wurm et al., 2015), kinematics (Wurm & Lingnau, 2015), and objects (Wurm & Lingnau, 2015) in LOTC, but not in PMv.

At first, these results seem difficult to reconcile with the studies described reporting PMv being activated during action observation, but they can be explained in terms of the type of action representation investigated. For example, the cross-modal action information found in the study by Kilner et al. (2009) was related to observed actions presented in first-person perspective, but this result does not hold anymore when one tries to generalize across perspectives (Oosterhof et al., 2012a). Similarly, we did find abstract action information in motor-related regions in the first MEG experiment (**Study 1**), in which we presented actions in first-person perspective, but not in the second study (**Study 2**), in which the actions were mostly presented in third-person perspectives. Moreover, results of **Study 1** can be explained following Wurm's findings (Wurm & Lingnau, 2015) that reported action discrimination in PMv only when the object involved in the action remained the same, as it was the case in our studies. Finally, the familiarity with certain types of actions could determine the involvement of

motor-related regions, as stated by the association hypothesis of mirror neurons (Heyes, 2010). The actions we used in **Study 1** were highly familiar (grasping/pointing) and might have a *special* status in motor-related regions. Further investigation on this issue is needed to disentangle between these three aspects.

To summarize, there is sufficient experimental evidence to state that motor-related regions do not encode high (conceptual) abstract representations of actions, but only intermediate representational levels. These results conflict with motor theories of action recognition (Pulvermüller, 2005; Rizzolatti et al., 2001). Our results are instead compatible with cognitive theories (Caramazza et al., 2014; Mahon & Caramazza, 2008), which claim that action recognition occurs in nonmotor regions that form the semantic system in which action concepts are located. Activation of the motor system during conceptual processing might be due to the fact that the systems are highly inter-connected. Therefore, motor-related activity might be post-conceptual and *reflect* action understanding (Csibra, 2007). If this is the case, where does action understanding occur?

Previous studies have shown that LOTC (**Figure 5.1**) is a well-suited region for hosting action information and related knowledge (for a review of the role of LOTC in action, see Lingnau & Downing, 2015). For example, fMRI studies reported that this region has a preferential response for hand-type stimuli (Bracci et al., 2010) and for body parts in general (Downing et al., 2001). LOTC shows stronger involvement when participants are required to discriminate between action types (e.g. punching vs kicking) than when they have to discriminate between effectors (Lingnau & Petris, 2013). Sources observed with MEG have also shown that LOTC is modulated during observation of point-light display (PLD) stimuli (Pavlidou et al., 2014a, 2014b). Moreover, this region is known to be recruited during the semantic processing of verbs (e.g. Papeo et al., 2014), and lesions to this region are associated with impairments in the ability to recognize actions (Kalénine et al., 2010; Urgesi et al., 2014). In line with this view, a recent lesion study including 139 patients which required matching a written phrase describing an everyday action to one of two actions presented via video (Tarhan et al.,

2015) demonstrated that lesions to primary motor, somatosensory and the inferior parietal lobule were accompanied by impaired action performance, whereas lesions in the posterior temporal region resulted in impaired action recognition. Furthermore, this region encodes abstract action information in monkeys (Jellema et al., 2000). The results reported in this thesis are perfectly in line with these findings and expand previous studies by adding the temporal dimension. Taken together, these studies suggest that LOTC is well suited to integrate various sources of information that are crucial for action understanding.

Future directions

As I have argued in this thesis, MEG in combination with MVPA is a suitable tool for investigating the neural bases underlying action representations. This technique is optimal in terms of temporal resolution and thus complementary to fMRI, which provides optimal spatial resolution. Together, these two neuroimaging methods can provide a strong contribution to this field. For example, an obvious follow up on my results would be to use fMRI to define the regions involved in the abstract encoding of action information. With MEG, I could determine *when* the action representation was available, but only provide an imprecise neural location. Knowing how information is spatially organized in a certain area, as LOTC, is determinant for building a model of action understanding that is biologically plausible.

Many other questions remain unanswered. *When* does action understanding actually occur? What additional aspects of an action are encoded in LOTC? What are the functional roles of motor-related regions in action observation?

To answer to these questions, it would be important to have a direct measure of action understanding. For example, one could design a study that investigates what happens in the instant *before* action understanding. This would help us to understand which sources are involved in building a percept of the observed action. MEG would be an ideal technique to investigate this issue. Finally, representational similarity analysis, or

RSA (Kriegeskorte et al., 2008), of fMRI and MEG data could help characterizing what type of action information is actually encoded in temporal regions.

Conclusions

The work presented in this thesis along with previous neuroimaging studies cast strong doubts on the claims that frontal motor-related regions, such as PMv, contain high-level (e.g. conceptual) abstract action representations. Furthermore, intermediate levels of action representation (e.g. within effector) are available with PMv only at a late latency. This region is part of a brain network typically involved in action observation (the action observation network) and has been claimed to play a causal role in action understanding. However, a region involved in action understanding should be capable of discriminating between actions early, in particular at high levels of abstraction. The results presented in my thesis provide evidence that LOTC can discriminate between abstract action representations at an early latency, supporting the view that this area plays a crucial role for action understanding.

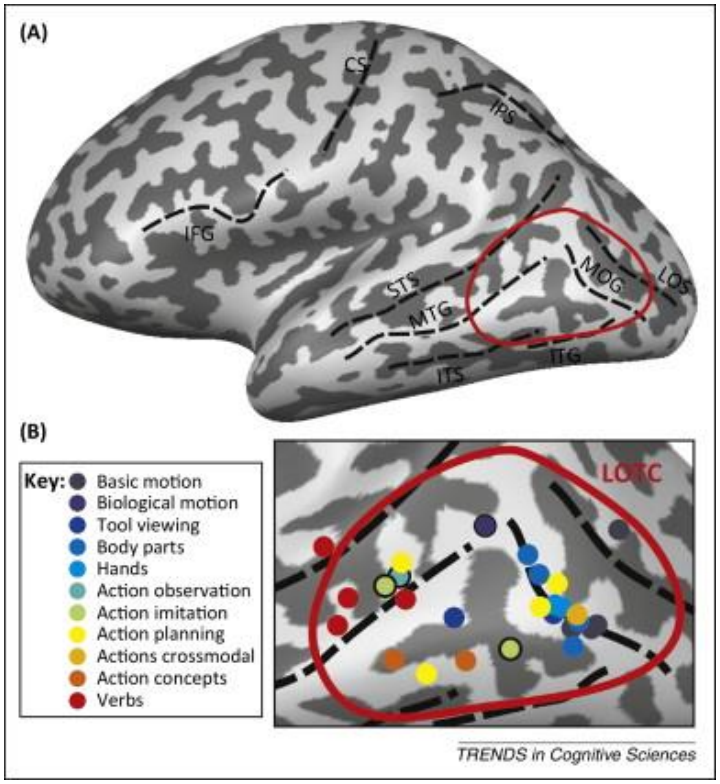
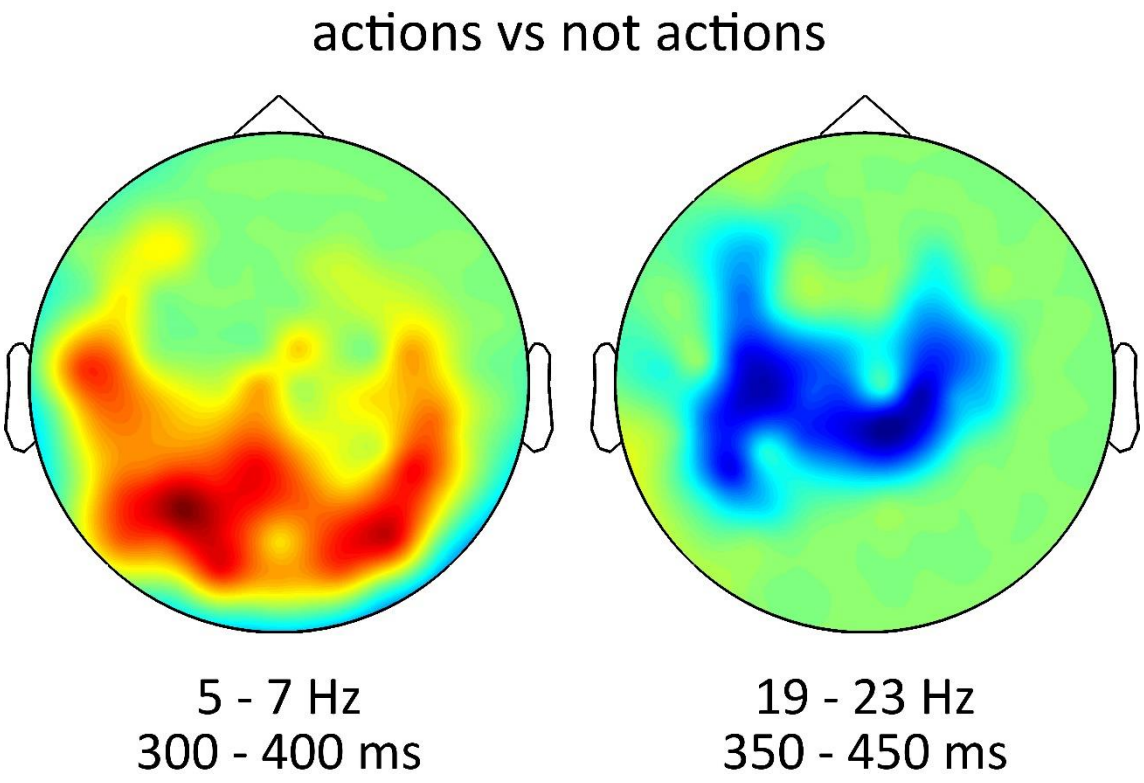


Figure 5.1. The lateral occipitotemporal cortex (LOTC). The red line in panel **A** and **B** indicates the approximate border of LOTC. Panel **B** shows also peak activations (coloured dots) of various studies that reported involvement of LOTC in different domains. Figure taken from Lingnau and Downing (2015).

Appendix A

Actions vs bodies

To evaluate if the motor system is involved during action observation in a generic and unspecific fashion, we conducted a univariate contrast between actions and bodies that we assumed did not contain any implicit movement and thus we considered as stimuli with no action information. In this sense, we defined the contrast as actions vs non-actions. We observed significant ($p<0.01$) differential modulation at low frequencies, with a peak in the theta-band at around 350 milliseconds, and at higher frequencies with a peak in beta-bands at around 400 milliseconds.



References

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–45. doi:10.1146/annurev.psych.59.103006.093639

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex (New York, N.Y. : 1991)*, *19*(12), 2767–96. doi:10.1093/cercor/bhp055

Blakemore, S.-J., & Frith, C. (2005). The role of motor contagion in the prediction of action. *Neuropsychologia*, *43*(2), 260–267. doi:10.1016/j.neuropsychologia.2004.11.012

Boulenger, V., Roy, A. C., Paulignan, Y., Deprez, V., Jeannerod, M., & Nazir, T. a. (2006). Cross-talk between language processes and overt motor behavior in the first 200 msec of processing. *Journal of Cognitive Neuroscience*, *18*(10), 1607–15. doi:10.1162/jocn.2006.18.10.1607

Bracci, S., Ietswaart, M., Peelen, M. V, & Cavina-Pratesi, C. (2010). Dissociable neural responses to hands and non-hand body parts in human left extrastriate visual cortex. *Journal of Neurophysiology*, *103*(6), 3389–97. doi:10.1152/jn.00215.2010

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. doi:10.1163/156856897X00357

Broca, P. (1861). Perte de la parole, ramollissement chronique et destruction partielle du lobe antérieur gauche du cerveau. *Bulletin de La Société Anthropologique*, *2*(1), 235–238.

Buxbaum, L. J., & Kalénine, S. (2010). Action knowledge, visuomotor activation, and embodiment in the two action systems. *Annals of the New York Academy of Sciences*, *1191*, 201–18. doi:10.1111/j.1749-6632.2010.05447.x

Buxbaum, L. J., Shapiro, A. D., & Coslett, H. B. (2014). Critical brain regions for tool-related and imitative actions: a componential analysis. *Brain : A Journal of Neurology*, *137*(Pt 7), 1971–85. doi:10.1093/brain/awu111

Caggiano, V., Giese, M., Thier, P., & Casile, A. (2014). Encoding of point of view during action observation in the local field potentials of macaque area F5. *The European Journal of Neuroscience*, (October), 1–11. doi:10.1111/ejn.12793

Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: an FMRI study with expert dancers. *Cerebral Cortex*, *15*(8), 1243–9. doi:10.1093/cercor/bhi007

Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology*,

References

16(19), 1905–10. doi:10.1016/j.cub.2006.07.065

Caramazza, A., Anzellotti, S., Strnad, L., & Lingnau, A. (2014). Embodied Cognition and Mirror Neurons: A Critical Assessment. *Annual Review of Neuroscience*, *37*, 1–15. doi:10.1146/annurev-neuro-071013-013950

Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *NeuroImage*, *50*(3), 1148–67. doi:10.1016/j.neuroimage.2009.12.112

Catmur, C., Walsh, V., & Heyes, C. (2007). Sensorimotor learning configures the human mirror system. *Current Biology*, *17*(17), 1527–31. doi:10.1016/j.cub.2007.08.006

Cattaneo, L., Sandrini, M., & Schwarzbach, J. (2010). State-dependent TMS reveals a hierarchical representation of observed acts in the temporal, parietal, and premotor cortices. *Cerebral Cortex (New York, N.Y. : 1991)*, *20*(9), 2252–8. doi:10.1093/cercor/bhp291

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 27:1–27:27.

Cochin, S., Barthelemy, C., Roux, S., & Martineau, J. (1999). Observation and execution of movement: similarities demonstrated by quantified electroencephalography. *European Journal of Neuroscience*, *11*(5), 1839–1842. doi:10.1046/j.1460-9568.1999.00598.x

Cross, E. S., Hamilton, A. F. D. C., & Grafton, S. T. (2006). Building a motor simulation de novo: Observation of dance by dancers. *NeuroImage*, *31*(3), 1257–1267. doi:10.1016/j.neuroimage.2006.01.033

Csibra, G. (2007). Action mirroring and action understanding: an alternative account. In *Sensorimotor Foundations of Higher Cognition*.

Decety, J., Grèzes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., … Fazio, F. (1997). Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain : A Journal of Neurology*, *120*(10), 1763–77. doi:10.1093/brain/120.10.1763

di Pellegrino, G., Fadiga, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *91*, 176–180.

Dinstein, I., Thomas, C., Behrmann, M., & Heeger, D. J. (2008). A mirror up to nature. *Current Biology*, *18*, R13–R18.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science (New York, N.Y.)*, *293*(5539), 2470–3. doi:10.1126/science.1063414

Downing, P. E., & Peelen, M. V. (2011). The role of occipitotemporal body-selective regions in person perception. *Cognitive Neuroscience*, *2*(3-4), 186–203.

doi:10.1080/17588928.2011.582945

Engel, A. K., Maye, A., Kurthen, M., & König, P. (2013). Where's the action? The pragmatic turn in cognitive science. *Trends in Cognitive Sciences*, *17*(5). doi:10.1016/j.tics.2013.03.006

Fabbri-Destro, M., & Rizzolatti, G. (2008). Mirror neurons and mirror systems in monkeys and humans. *Physiology (Bethesda, Md.)*, *23*, 171–9. doi:10.1152/physiol.00004.2008

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic stimulation study. *Journal of Neurophysiology*, *73*(6), 2608–2611.

Fairhall, S. L., & Caramazza, A. (2013). Brain Regions That Represent Amodal Conceptual Knowledge. *Journal of Neuroscience*, *33*(25), 10552–10558. doi:10.1523/JNEUROSCI.0051-13.2013

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science (New York, N.Y.)*, *308*(5722), 662–7. doi:10.1126/science.1106138

Frenkel-Toledo, S., Bentin, S., Perry, A., Liebermann, D. G., & Soroker, N. (2013). Dynamics of the EEG power in the frequency and spatial domains during observation and execution of manual movements. *Brain Research*, *1509*, 43–57. doi:10.1016/j.brainres.2013.03.004

Frith, C. D., & Frith, U. (1999). Interacting Minds — A Biological Basis. *Science*, *286*(November), 1692–1695. doi:10.1126/science.286.5445.1692

Gallese, V., Fadiga, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain : A Journal of Neurology*, *119 (Pt 2*, 593–609.

Gazzola, V., & Keysers, C. (2009). The observation and execution of actions share motor and somatosensory voxels in all tested subjects: single-subject analyses of unsmoothed fMRI data. *Cerebral Cortex*, *19*(6), 1239–55. doi:10.1093/cercor/bhn181

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–65.

Grafton, S. T., & Hamilton, A. F. D. C. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human Movement Science*, *26*(4), 590–616. doi:10.1016/j.humov.2007.05.009

Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *PNAS*, *98*(2), 694–9. doi:10.1073/pnas.98.2.694

Hamilton, A. F. D. C. (2009). Research review: Goals, intentions and mental states: challenges for theories of autism. *Journal of Child Psychology and Psychiatry*, *50*(8), 881–892. doi:10.1111/j.1469-7610.2009.02098.x

Hamilton, A. F. D. C., & Grafton, S. T. (2006). Goal representation in human anterior

intraparietal sulcus. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *26*(4), 1133–7. doi:10.1523/JNEUROSCI.4551-05.2006

Hamilton, A. F. D. C., & Grafton, S. T. (2008). Action outcomes are represented in human inferior frontoparietal cortex. *Cerebral Cortex (New York, N.Y. : 1991)*, *18*(5), 1160–8. doi:10.1093/cercor/bhm150

Hari, R. (2006). Action-perception connection and the cortical mu rhythm. *Progress in Brain Research*, *159*(06), 253–60. doi:10.1016/S0079-6123(06)59017-X

Hauk, O., & Pulvermüller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Human Brain Mapping*, *21*(3), 191–201. doi:10.1002/hbm.10157

Hauk, O., Shtyrov, Y., & Pulvermüller, F. (2008). The time course of action and action-word comprehension in the human brain as revealed by neurophysiology. *Journal Of Physiology Paris*, *102*(1-3), 50–8. doi:10.1016/j.jphysparis.2008.03.013

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: The early beginnings. *NeuroImage*, 1–4. doi:10.1016/j.neuroimage.2012.03.016

Haxby, J. V, Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science (New York, N.Y.)*, *293*(5539), 2425–30. doi:10.1126/science.1063736

Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience and Biobehavioral Reviews*, *34*(4), 575–83. doi:10.1016/j.neubiorev.2009.11.007

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience*, *21*(7), 1229–43. doi:10.1162/jocn.2009.21189

Hickok, G. (2012). Do mirror neurons subserve action understanding? *Neuroscience Letters*, *540*, 56–58. doi:10.1016/j.neulet.2012.11.001

Hickok, G., & Hauser, M. (2010). (Mis)understanding mirror neurons. *Current Biology*, *20*(14), R593–R594. doi:10.1016/j.cub.2010.05.047

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, *148*(3), 574–591.

Jellema, T., Baker, C. I., Wicker, B., & Perrett, D. I. (2000). Neural Representation for the Perception of the Intentionality of Actions. *Brain and Cognition*, *44*(2), 280–302. doi:10.1006/brcg.2000.1231

Jensen, O., & Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *European Journal of Neuroscience*, *15*(8), 1395–1399. doi:10.1046/j.1460-9568.2002.01975.x

Kalénine, S., Buxbaum, L. J., & Coslett, H. B. (2010). Critical brain regions for action recognition:

lesion symptom mapping in left hemisphere stroke. *Brain : A Journal of Neurology*, *133*(11), 3269–80. doi:10.1093/brain/awq210

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *48*(7), 805–25. doi:10.1016/j.cortex.2011.04.006

Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in Cognitive Sciences*, *15*(8), 352–7. doi:10.1016/j.tics.2011.06.005

Kilner, J. M., Kraskov, A., & Lemon, R. N. (2014). Do monkey F5 mirror neurons show changes in firing rate during repeated observation of natural actions ? *Journal of Neurophysiology*, *111*(December 2013), 1214–1226. doi:10.1152/jn.01102.2012

Kilner, J. M., Neal, A., Weiskopf, N., Friston, K. J., & Frith, C. D. (2009). Evidence of mirror neurons in human inferior frontal gyrus. *The Journal of Neuroscience*, *29*(32), 10153–9. doi:10.1523/JNEUROSCI.2668-09.2009

Kilner, J. M., Paulignan, Y., & Blakemore, S. J. (2003). An interference effect of observed biological movement on action. *Current Biology : CB*, *13*(6), 522–5. doi:10.1016/S0960-9822(03)00165-9

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. doi:10.1016/j.tics.2014.01.002

Koelewijn, T., van Schie, H. T., Bekkering, H., Oostenveld, R., & Jensen, O. (2008). Motor-cortical beta oscillations are modulated by correctness of observed action. *NeuroImage*, *40*(2), 767–75. doi:10.1016/j.neuroimage.2007.12.018

Kokal, I., Gazzola, V., & Keysers, C. (2009). Acting together in and beyond the mirror neuron system. *NeuroImage*, *47*(4), 2046–56. doi:10.1016/j.neuroimage.2009.06.010

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *PNAS*, *103*(10), 3863–8. doi:10.1073/pnas.0600244103

Kriegeskorte, N., Mur, M., & Bandettini, P. a. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(November), 4. doi:10.3389/neuro.06.004.2008

Leshinskaya, A., & Caramazza, A. (2015). Abstract categories of functions in anterior parietal lobe. *Neuropsychologia*, 1–13. doi:10.1016/j.neuropsychologia.2015.01.014

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461. doi:10.1037/h0020279

Lingnau, A., & Downing, P. E. (2015). The lateral occipitotemporal cortex in action. *Trends in Cognitive Sciences*. doi:10.1016/j.tics.2015.03.006

References

Lingnau, A., & Petris, S. (2013). Action understanding within and outside the motor system: the role of task difficulty. *Cerebral Cortex (New York, N.Y. : 1991)*, *23*(6), 1342–50. doi:10.1093/cercor/bhs112

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology, Paris*, *102*(1-3), 59–70. doi:10.1016/j.jphysparis.2008.03.004

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–90. doi:10.1016/j.jneumeth.2007.03.024

Mur, M., Bandettini, P. a, & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI--an introductory guide. *Social Cognitive and Affective Neuroscience*, *4*(1), 101–9. doi:10.1093/scan/nsn044

Muthukumaraswamy, S. D., & Johnson, B. W. (2004). Changes in rolandic mu rhythm during observation of a precision grip. *Psychophysiology*, *41*(1), 152–6. doi:10.1046/j.1469-8986.2003.00129.x

Muthukumaraswamy, S. D., Johnson, B. W., & McNair, N. a. (2004). Mu rhythm modulation during observation of an object-directed grasp. *Brain Research. Cognitive Brain Research*, *19*(2), 195–201. doi:10.1016/j.cogbrainres.2003.12.001

Negri, G. A. L., Rumiati, R. I., Zadini, A., Ukmar, M., Mahon, B. Z., & Caramazza, A. (2007). What is the role of motor simulation in action and object recognition? Evidence from apraxia. *Cognitive Neuropsychology*, *24*(8), 795–816. doi:10.1080/02643290701707412

Nelissen, K., Borra, E., Gerbella, M., Rozzi, S., Luppino, G., Vanduffel, W., … Orban, G. a. (2011). Action observation circuits in the macaque monkey cortex. *The Journal of Neuroscience*, *31*(10), 3743–56. doi:10.1523/JNEUROSCI.4803-10.2011

Nishitani, N., & Hari, R. (2000). Temporal dynamics of cortical representation for action. *PNAS*, *97*(2), 913–8.

Nishitani, N., & Hari, R. (2002). Viewing Lip Forms. *Neuron*, *36*(6), 1211–1220. doi:10.1016/S0896-6273(02)01089-9

Nolte, G. (2003). The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Physics in Medicine and Biology*, *48*(22), 3637–3652. doi:10.1088/0031-9155/48/22/002

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. doi:10.1155/2011/156869

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (in preparation). CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab / GNU Octave. Toolbox available from http://cosmomvpa.org.

References

Oosterhof, N. N., Tipper, S. P., & Downing, P. E. (2012a). Viewpoint (in)dependence of action representations: an MVPA study. *Journal of Cognitive Neuroscience*, *24*(4), 975–89. doi:10.1162/jocn_a_00195

Oosterhof, N. N., Tipper, S. P., & Downing, P. E. (2012b). Visuo-motor imagery of specific manual actions: a multi-variate pattern analysis fMRI study. *NeuroImage*, *63*, 262–271.

Oosterhof, N. N., Tipper, S. P., & Downing, P. E. (2013). Crossmodal and action-specific: neuroimaging the human mirror neuron system. *Trends in Cognitive Sciences*, *17*(7), 311–8. doi:10.1016/j.tics.2013.04.012

Oosterhof, N. N., Wiggett, A. J., Diedrichsen, J., Tipper, S. P., & Downing, P. E. (2010). Surface-based information mapping reveals crossmodal vision-action representations in human parietal and occipitotemporal cortex. *Journal of Neurophysiology*, *104*(2), 1077–89. doi:10.1152/jn.00326.2010

Papeo, L., Lingnau, A., Agosta, S., Pascual-Leone, A., Battelli, L., & Caramazza, A. (2014). The Origin of Word-related Motor Activity. *Cerebral Cortex (New York, N.Y. : 1991)*, 1–8. doi:10.1093/cercor/bht423

Pavlidou, A., Schnitzler, A., & Lange, J. (2014a). Distinct spatio-temporal profiles of beta-oscillations within visual and sensorimotor areas during action recognition as revealed by MEG. *Cortex*, *54*, 106–16. doi:10.1016/j.cortex.2014.02.007

Pavlidou, A., Schnitzler, A., & Lange, J. (2014b). Interactions between visual and motor areas during the recognition of plausible actions as revealed by magnetoencephalography. *Human Brain Mapping*, *35*(2), 581–592. doi:10.1002/hbm.22207

Pazzaglia, M., Smania, N., Corato, E., & Aglioti, S. M. (2008). Neural Underpinnings of Gesture Discrimination in Patients with Limb Apraxia. *Journal of Neuroscience*, *28*(12), 3030–3041. doi:10.1523/JNEUROSCI.5748-07.2008

Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain : A Journal of Neurology*, *60*(4), 389–443.

Pereira, F., Mitchell, T. M., & Botvinick, M. M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage*, *45*(1 Suppl), S199–209. doi:10.1016/j.neuroimage.2008.11.007

Perrett, D. I., Harries, M., Bevan, R., Thomas, S., Benson, P., Mistlin, A., … Ortega, J. (1989). Frameworks of analysis for the neural representation of animate objects and actions. *Journal of Experimental Biology*, *113*, 87–113.

Pfurtscheller, G., & Lopes da Silva, F. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, *110*(11), 1842–57.

Pineda, J. a. (2005). The functional significance of mu rhythms: translating “seeing” and

References

"hearing" into "doing". *Brain Research. Brain Research Reviews*, *50*(1), 57–68. doi:10.1016/j.brainresrev.2005.04.005

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews. Neuroscience*, *6*(7), 576–82. doi:10.1038/nrn1706

Pulvermüller, F. (2013). Semantic embodiment, disembodiment or misembodiment? In search of meaning in modules and neuron circuits. *Brain and Language*, *127*(1), 86–103. doi:10.1016/j.bandl.2013.05.015

Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience*, *17*(6), 884–92. doi:10.1162/0898929054021111

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. *Experimental Brain Research*, *71*(3), 491–507. doi:10.1007/BF00248742

Rizzolatti, G., Cattaneo, L., Fabbri-Destro, M., & Rozzi, S. (2014). Cortical mechanisms underlying the organization of goal-directed actions and mirror neuron-based action understanding. *Physiological Reviews*, *94*(2), 655–706. doi:10.1152/physrev.00009.2013

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–92. doi:10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., Fadiga, L., & Gallese, V. (1996). Premotor cortex and the recognition of motor actions. *Brain Research. Cognitive Brain Research*, *3*(2), 131–41.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews. Neuroscience*, *2*(9), 661–70. doi:10.1038/35090060

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature Reviews. Neuroscience*, *11*(4), 264–74. doi:10.1038/nrn2805

Schwarzbach, J. (2011). A simple framework (ASF) for behavioral and neuroimaging experiments based on the psychophysics toolbox for MATLAB. *Behavior Research Methods*, *43*(4), 1194–201. doi:10.3758/s13428-011-0106-8

Shmuelof, L., & Zohary, E. (2007). Watching others' actions: mirror representations in the parietal cortex. *The Neuroscientist : A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *13*(6), 667–72. doi:10.1177/1073858407302457

Tarhan, L. Y., Watson, C. E., & Buxbaum, L. J. (2015). Shared and Distinct Neuroanatomic Regions Critical for Tool-related Action Production and Recognition: Evidence from 131 Left-hemisphere Stroke Patients. *Journal of Cognitive Neuroscience*, 1–21. doi:10.1162/jocn_a_00876

References

Turella, L., Wurm, M. F., Tucciarelli, R., & Lingnau, A. (2013). Expertise in action observation: recent neuroimaging findings and future perspectives. *Frontiers in Human Neuroscience*, *7*(October), 637. doi:10.3389/fnhum.2013.00637

Turken, A. U., & Dronkers, N. F. (2011). The neural architecture of the language comprehension network: converging evidence from lesion and connectivity analyses. *Frontiers in Systems Neuroscience*, *5*, 1. doi:10.3389/fnsys.2011.00001

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., … Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273–89. doi:10.1006/nimg.2001.0978

Umiltà, M. A., Kohler, E., Gallese, V., Fadiga, L., Keysers, C., & Rizzolatti, G. (2001). I know what you are doing. a neurophysiological study. *Neuron*, *31*(1), 155–65.

Urgesi, C., Candidi, M., & Avenanti, A. (2014). Neuroanatomical substrates of action perception and understanding: an anatomic likelihood estimation meta-analysis of lesion-symptom mapping studies in brain injured patients. *Frontiers in Human Neuroscience*, *8*, 344. doi:10.3389/fnhum.2014.00344

Valyear, K. F., & Culham, J. C. (2010). Observing learned object-specific functional grasps preferentially activates the ventral stream. *Journal of Cognitive Neuroscience*, *22*(5), 970–84. doi:10.1162/jocn.2009.21256

Williams, J. H. G., Whiten, A., Suddendorf, T., & Perrett, D. I. (2001). Imitation, mirror neurons and autism. *Neuroscience & Biobehavioral Reviews*, *25*(4), 287–295. doi:10.1016/S0149-7634(01)00014-8

Wurm, M. F., Ariani, G., Greenlee, M. W., & Lingnau, A. (2015). Decoding Concrete and Abstract Action Representations During Explicit and Implicit Conceptual Processing. *Cerebral Cortex*, bhv169. doi:10.1093/cercor/bhv169

Wurm, M. F., & Lingnau, A. (2015). Decoding Actions at Different Levels of Abstraction. *Journal of Neuroscience*, *35*(20), 7727–7735. doi:10.1523/JNEUROSCI.0188-15.2015